International Journal of Information Security (2025) 24:86
https://doi.org/10.1007/s10207-025-00999-0

REGULAR CONTRIBUTION

The SPADA methodology for threat modelling

Mario Raciti[1,2] · **Giampaolo Bella**[2]

Published online: 1 March 2025

Abstract

As individuals engage with innovative technologies, including smart cars and smart homes, a comprehensive treatment of the threats to their privacy becomes increasingly urgent. This article recognises the relevance of security and, in particular, privacy threat modelling, especially under the umbrella of GDPR compliance, and addresses the challenge of the pursuit of completeness in eliciting security and privacy threats. The core contribution is SPADA, a methodology for threat modelling revolving around five key variables (whose initials form the acronym that names the methodology). These are: "Source of documentation", "Property", "Application domain", "Detail (level of)" and "Agent(s) raising the threats", and clarify the essential variable elements of the threat modelling activity. SPADA requires the analyst to duly instantiate each variable but offers increased structure and automation in return. The methodology is applied to the domains of smart cars and smart homes, considering both soft and hard privacy. This yields 23 domain-independent threats for soft privacy, and 29 domain-independent threats for hard privacy. Both these lists of threats are then tailored to the smart car domain by appropriate combination with the 43 identified assets, producing a total of 785 privacy threats for smart cars. Similarly, appropriate combination with the 127 assets identified in the smart home domain produces a total of 1502 privacy threats for smart homes.

Keywords Risk assessment · Smart car · Smart home · ENISA · LINDDUN

1 Introduction

The rapid adoption of technology in modern society has given rise to an intricate network of devices, each communicating and sharing data to facilitate seamless user experiences. While offering unprecedented convenience, this pervasive interconnectivity has also magnified the potential for privacy breaches. As a consequence, it has become essential to identify, understand, and ultimately mitigate the risks associated with data exposure and privacy infringements. Within this landscape, from connected vehicles transmitting geolocation information to service providers offering personalised recommendations, the ingredients required to prepare a perfect recipe for a comprehensive privacy threat modelling need additional scrutiny. Privacy is a complex and multifaceted concept that may be interpreted in different ways in different contexts, yet in the first place it is considered a fundamental human right by many. NIST defined it as "*the right of a party to maintain control over and confidentiality of information about itself*" [41]. In a General Data Protection Regulation (GDPR) fashion, we may summarise privacy as the right of an individual, i.e., the data subject, to control or influence what information related to them may be collected, processed and stored, and by whom and to whom that information may be disclosed.

Hard privacy is based on an interpretation of privacy that is tied to the confidentiality of a subject's personal data [14, 29]. Briefly, hard privacy concerns the various techniques to protect a subject's personal data from everyone else, hence it calls for measures such as anonymisation and minimisation. By contrast, *soft privacy* represents the ability to control what happens with a data subject's personal data [14, 29]. Consequently, soft privacy pertains to the range of practices to be followed for the subject to share their personal data with someone else while keeping full control, thus measures such as consent mechanisms and impact assessments are suit-

Mario Raciti and Giampaolo Bella have contributed equally to this work.

✉ Mario Raciti
mario.raciti@imtlucca.it

Giampaolo Bella
giampaolo.bella@unict.it

[1] IMT School for Advanced Studies Lucca, Piazza S. Ponziano, 6, 55100 Lucca, Italy

[2] Università degli Studi di Catania, Viale Andrea Doria, 6, 95125 Catania, Italy

able. The pursuit of completeness in eliciting (hard and soft) privacy threats necessitates guidelines to steer the analyst through the process of threat elicitation.

Furthermore, the imperative for a complete enumeration of privacy threats is reinforced by the backdrop of the GDPR's ascendancy as a global benchmark. The so-called "Brussels effect" is in full force, orienting nations, industries, and researchers to elevate their privacy governance paradigms to meet or exceed the stringent benchmarks set forth by the GDPR. As this effect pervades the international privacy discourse, the quest for a comprehensive list of privacy threats gains added resonance.

Moreover, the 2024 Global Automotive Consumer Study by Deloitte [15] confirms the awareness raised by the GDPR in the particular domain of smart cars. The study reports that trust issues are hindering consumers, especially in Europe, from sharing their personal data. This also witnesses that Smart Cars [19], Smart Roads [47] and Smart Cities [58] have gained significant attention, as they generate vast amounts of data that require secure storage, transmission, and processing. Additionally, the integration of various sensors, cameras, and communication systems in modern vehicles creates new opportunities for privacy breaches, raising concerns about data protection measures and corresponding risks.

Nevertheless, the smart car domain is not the sole area drawing attention from a privacy perspective. As a parallel subset of the IoT domain, smart homes hold considerable implications for individual privacy across various dimensions. These may encompass personal information, behavioural patterns, communication channels, data integrity, visual content, geographical particulars, and associations. In the trajectory towards pervasive computing, the role of privacy takes on heightened significance. Within smart home systems, concealed functionalities may operate beyond user awareness, omitting essential updates and transparency, thus raising privacy concerns [8]. Notably, smart homes can incorporate sensitive components tied to inhabitants' well-being, financial transactions, and mechanisms pivotal to household security-areas that potentially invite malicious manipulation from (external) adversaries.

Our research rests on the observation that the process of modelling privacy threats has not obtained the same attention as the traditional, cybersecurity threat modelling so far. Moreover, at least GDPR compliance demands a privacy risk assessment, which in turn demands privacy threat modelling, hence the general motivation for this article.

1.1 Context

Privacy is frequently intertwined with security, as privacy concerns frequently arise in relation to security matters. While, as we shall see below, threat modelling has traditionally been tackled from a security perspective, it should be emphasised that privacy and security are two distinct concepts, hence they cannot be used interchangeably. A challenge for privacy threat modelling in general is how to consider the impact on data subjects involved in the privacy threat. This aspect is stressed in law and regulations compliance, e.g., in the Data Protection Impact Assessment (DPIA), required under the GDPR, to help identify, assess, and mitigate privacy risks associated with data processing activities. From a data subject's perspective, a DPIA is an important aspect of privacy threat modelling. It ensures that organisations consider the potential impact of their data processing on individuals' privacy rights and take appropriate measures to address any risks. Obviously, a DPIA would benefit from a privacy threat model.

Furthermore, threat modelling is challenging as the analyst faces various problems, such as completeness and threat explosion. On the one hand, completeness may be impactful because failing to account for specific threats would cause pitfalls to the subsequent risk assessment. On the other hand, the pursuit of completeness can result in a phenomenon known as threat explosion, characterised by an overwhelming number of threats that may be irrelevant, infeasible, or redundant with each other. Completeness and redundancy are considered by our previous work that features *threat embracing* [49]. Briefly, if two or more threats descriptions are deemed redundant in terms of their semantic similarity by the analyst's scrutiny, then these threats can be semantically merged into one.

In addition, as we shall see below, while security threats to smart cars and smart homes have been widely analysed, there is a lack of privacy threat taxonomies that focus on a comprehensive treatment of both hard and soft privacy for smart cars and smart homes in the state of the art, hence a clear motivation to push towards the advancement of a security and privacy threat modelling framework that can be specifically tailored to both domains. We built an early version of soft privacy threats for smart cars by taking a domain-dependent approach and by leveraging the threats from various sources [50]. These included the LINDDUN state-of-the-art privacy threat modelling framework [60] and ENISA's "Good practices for security of smart cars" [19]. In particular, although ENISA's report is among the most relevant sources about car cybersecurity in Europe, its treatment of privacy is very limited, hence the need for a deeper close-up. Moreover, the situation for smart homes is even more challenging. Unlike cybersecurity, where frameworks and taxonomies have gained significant traction also for the smart home domain, this remains relatively underrepresented in terms of dedicated privacy threat models and privacy threat knowledge. Also in this case, ENISA provides a report, i.e., "Threat Landscape and Good Practice Guide for Smart Home and Converged Media" [17], with a rather scarce treatment of privacy.

Another element of context is that LINDDUN has recently been significantly updated, hence the results from our previous work for the smart car domain demand accurate revision. More precisely, LINDDUN has raised the number of soft privacy threats and, in consequence, we started to work an up-to-date list of soft privacy threats for the smart car domain [51]. This could be pursued by leveraging the new version of the LINDDUN methodology, and it would bring the useful by-product of checking how LINDDUN has evolved over time, particularly whether in the same direction we advocated [49].

Remarkably, hard privacy has not received the same attention as soft privacy so far, hence the motivation to apply the same methodology to provide an exhaustive soft-and-hard-privacy threat taxonomy for the smart car and smart home domains. This would strengthen the baseline for a threat modelling framework that is general for security and privacy and that can be tailored in the future to specific application domains.

1.2 Research question

Following the motivations and context given above, this article focuses on both soft and hard privacy from the threat modelling perspective. With the aim of advancing previous research, this article addresses the core research question:

RQ *How to model (soft and hard) privacy threats?*

To go about such a question, we observe that the mentioned LINDDUN methodology is widely established [16]. So, an answer could be found, potentially, in such a methodology. However, LINDDUN has two major drawbacks. The first is that it is meant to be domain-independent, a feature that is bound to keep its threat descriptions only at an abstract level of detail. The second is that the LINDDUN threat knowledge base may not be comprehensive and exhaustive, as demonstrated by our previous work [49] for soft privacy. Moreover, we question whether we can obtain a comprehensive threat knowledge base for two timely and relevant domains, i.e., smart car and smart home, hence we set the following specific research questions:

SRQ1. *What are the soft and hard privacy threats for the smart car domain?*

SRQ2. *What are the soft and hard privacy threats for the smart home domain?*

1.3 Contributions

This article answers the research questions by advancing the SPADA methodology for threat modelling, a refinement of the methodology proposed in our previous work [50], and by applying it to the current landscape of the smart car and smart home domains. By incorporating five variable elements into the analysis, the SPADA methodology ensures that the direction pursued by the analyst remains focused and aligned with the desired outcome. The variables act as guiding principles, allowing the analyst to make informed decisions based on relevant and reliable information. In addition, we apply the SPADA methodology for threat modelling to address the research questions. In particular, we propose two exercises for each domain (thus a total of four exercises) to focus on both soft privacy and hard privacy with the new version of LINDDUN. The contributions presented in this article are manyfold.

1.3.1 Extension of domain-independent threats

SPADA adopts, in particular, the mentioned ENISA reports on smart cars and smart homes as sources of specific and comprehensive knowledge on the target domains, and OWASP's "Calculation of the complete Privacy Risks list v2.0" [42]. These sources are augmented with the new version of LINDDUN and with an additional representative of the state of the art, i.e., the ENISA "Threat Taxonomy v2016" [18]. Therefore, SPADA rests on a significantly extended, domain-independent threat knowledge base. In particular, this article provides an updated list of 23 soft privacy threats that are domain-independent, thereby extending the 17 that we made available when we adopted the previous version of LINDDUN [50]. Because LINDDUN's soft privacy threats have increased from 9 to 17 over its two versions, our proposed extensions of it have decreased from 8 to 6. As we shall detail below, this can be taken as an indication that LINDDUN has evolved in the direction we advocated.

1.3.2 Soft privacy threats for smart car and smart home

Our novel 23 domain-independent soft privacy threats are also appropriately combined with 43, rather than 41 as before [50], specific assets of the smart car domain, so as to produce a total of 525 domain-dependent soft privacy threats for the smart car domain. In addition to the extended list of domain-dependent threats for smar cars, our novel, extended 23 domain-independent threats are appropriately combined with 127 specific assets of the smart home domain, thus producing a total of 1158 domain-dependent soft privacy threats for the smart home domain. Each combination instantiates a given threat to each of the assets that are deemed affected by the threat. These represent a substantial extension to the threat taxonomies existing in the state of the art, as multiple sources are combined, thereby supporting the argument that a better understanding of (soft) privacy within both the smart car and smart home domains is achieved.

1.3.3 Hard privacy threats for smart car and smart home

This article also provides, for the first time, a list of 29 hard privacy threats that are domain-independent. These threats are also appropriately combined with the 43 specific assets of the smart car domain, so as to produce a total of 260 domain-dependent hard privacy threats for the smart car domain. In addition, our novel 29 threats are appropriately combined with the 127 specific assets of the smart home domain, thereby producing a total of 344 domain-dependent hard privacy threats for the smart home domain. Also in this case, each combination instantiates a given threat to each of the assets that are deemed affected by the threat. Also for hard privacy, these represent a substantial extension to the threat taxonomies existing in the state of the art, thereby expanding the argument that a better understanding of (hard) privacy, hence privacy in its entirety, within both the smart car and smart home domains is achieved.

1.3.4 Repository of privacy threats

This article releases, for the first time, a repository available online [48] containing the extended domain-independent threat knowledge base for both soft and hard privacy. The repository includes the lists of assets collected in the smart car and smart home domains, along with domain-dependent threat knowledge bases for soft and hard privacy in these domains. Further detailed information is available in the repository readme file.

1.4 Article summary

Early components inherited by the SPADA methodology were introduced in two conference papers [50, 51]. The present article incorporates them, develops the full methodology and provides in-depth justification and detail. In particular, the final list of hard privacy threats for smart cars and both the lists of soft and hard privacy threats for smart homes are unpublished. The rest of the manuscript is organised as follows. Section 2 outlines the related work, and Sect. 3 gives an overview of LINDDUN and its latest changes. Section 4 provides the scientific reasoning behind the development of the SPADA methodology. Section 5 presents the variable elements of threat modelling. Section 6 describes the SPADA methodology for threat modelling. Section 7 demonstrates the methodology by applying it, respectively, to the smart car domain and smart home domain. Section 8 presents a partial validation to confirm the practicality and relevance of the results, and Sect. 9 evaluates the results. Section 10 concludes.

2 Related work

This Section conveniently treats related work in three categories, i.e., general, smart car, and smart home. As we shall see, the following works addressed crucial topics such as threat elicitation, threat knowledge base, security and privacy threat analysis and risk assessment, both in general and somewhat tailored to the smart car and smart home domains. However, to the best of our knowledge, there are no works advancing threat modelling considering its variable elements with the aim of comprehensively eliciting both domain-independent and domain-dependent, security, soft and hard privacy threats, for example upon the basis of the de-facto standard LINDDUN methodology in its new version. These are the distinctive features of the present contribution.

2.1 General

The challenges implicated by threat modelling led Wuyts et al. [66] to highlight the problems of current knowledge bases, such as limited semantics and lack of instantiating logic. Also, the authors discussed the requirements for a privacy threat knowledge base that streamlines threat elicitation efforts. Furthermore, it is also noteworthy to recall that the process of threat modelling inherently implies assumptions and arbitrary decisions. Landuyt et al. [60] highlighted the influence of assumptions to the outcomes of the analysis during the risk assessment process, more precisely in the threat modelling phase in the context of a LINDDUN privacy threat elicitation.

Moreover, threat elicitation and, more in general, threat modelling can support security and privacy requirements elicitation methods [1, 39], which are crucial for developing systems that incorporate security-by-design and privacy-by-design to protect users effectively. Several methodologies have been proposed to identify these requirements, with some integrating privacy principles [37]. While Pattakou et al. [44] and Canedo et al. [9] reviewed security and privacy requirement engineering and elicitation methods to highlight the necessity of integrating these requirements from the early stages of system design, Islam et al. [31] were among the first who attempted to leverage relevant laws and regulations as a source for the elicitation of security and privacy requirements. In particular, the authors highlighted the challenging deriving from concepts and terminology used for requirements engineering, which are mostly different to those used in the legal domain, and the lack of appropriate modelling languages and techniques to support such activities. Naturally, the same challenges reflect to the process of threat elicitation as well.

2.2 Smart car

In addition, several attempts were made for the purposes of threat modelling in the smart car domain. Vasenev et al. [61] were among the first to apply an extended version of STRIDE [38] and LINDDUN [16] to conduct a threat analysis on security and privacy threats in the smart car domain. In particular, the case study is specific to long term support scenarios for over-the-air updates, which means the threat analysis lacks of broader attention to the general picture in the smart car domain. Moreover, this work suggests that the privacy topic in the smart car domain has not reached the same level of maturity as cybersecurity.

In general, threat modelling is part of the wider process that is risk assessment. Wang et al. [62] proposed a threat-oriented risk assessment framework tailored for the smart car domain, with the aim, among the others, of overcoming assumptions and subjectivity. This framework can be considered a precursor to ISO/IEC:21434 [32], which was defined a year later. Also, the authors applied STRIDE and the attack tree method for the threat modelling. In addition, De Gusmão et al. [27] proposed a risk analysis framework that adopts fault tree analysis. However, in both cases the focus lies more on cybersecurity than privacy aspects.

Moreover, Chah et al. [10] applied the LINDDUN methodology to elicit and analyse privacy requirements of CAV system, while respecting the privacy properties set by the GDPR. Such attempt represents a solid baseline for the broader process of privacy risk assessment tailored for the smart car domain. Despite the application of LINDDUN, which tailored the analysis to privacy, the descriptions of the (limited and) available threats seem to predominantly focus on cybersecurity aspects or, at most, hard privacy. In addition, the CAV system only represents a subset of the broader smart car domain.

Finally, Bella et al. [4] advanced a dedicated risk assessment framework for privacy risks in smart cars. They proposed a double assessment, combining an asset-oriented ISO approach with a threat-oriented STRIDE approach. Also in this case, even though the treatment was fully tailored to the smart car domain, the privacy threats that were elicited appear to be more on cybersecurity aspects or, at most, hard privacy.

2.3 Smart home

For what concerns the smart home domain, Ghirardello et al. [25] introduced a reference architecture for smart homes through an exploration of three distinct perspectives within the ecosystem: (i) the functional perspective, encompassing essential operations required for the smart home's regular functioning; (ii) the physical perspective, detailing the physical elements crucial for executing the smart home's functions; and (iii) the communication perspective, outlining the essential protocols for transmitting control and information flows among these components. This reference architecture was then leveraged by Kavallieratos et al. [33], who examined existing dynamic risk assessment methodologies and identified security risks of a smart home's physical and communication viewpoints. Moreover, the relevance of IoT risks was analysed by Brous et al. [7], featuring a review focused on risks deriving from the adoption of IoT devices by organisations.

From a pure threat-elicitation point of view, Ziegeldorf et al. [67] were among the first to analyse privacy threats and challenges faced by the IoT. Geneiatakis et al. [24] proposed a security and privacy threat analysis for a typical smart home architecture, with a focus on the flaws introduced in smart homes through interactions among different devices. However, the majority of the elicited threats target cybersecurity aspects, with just two threats touching hard privacy. Furthermore, Heartfield et al. [26] classified cyber threats targeting smart homes according to a novel taxonomy that focuses not only on the attack vectors that can be used, but also on the potential impact on the systems and the occupants. Yet, the taxonomy lacks of privacy threats, rather including cybersecurity.

In addition, Siwakoti et al. [55] provided a review of recent advances in vulnerabilities, threats, and attacks in the most-generic field of IoT (thus covering both smart car and smart home, among the others sub-application domains), including a study on criminal services leveraging such elements. Finally, Anwar et al. [2] proposed a threat taxonomy for smart homes, yet limited to a cybersecurity perspective rather than privacy.

3 A primer on (the new) LINDDUN

It is convenient to provide an introduction to LINDDUN before proceeding with the description of the SPADA methodology for at least two reasons. The first is that LINDDUN is a de-facto standard privacy threat modelling methodology. The second is that SPADA leverages LINDDUN as one of the sources for threat knowledge base. Inspired by STRIDE, LINDDUN supports analysts in the systematical elicitation and mitigation of privacy threats in software architectures. LINDDUN privacy knowledge base represents one of its main strengths, and it is structured according to the seven privacy threat categories encapsulated within LINDDUN's acronym [16]. Recently, LINDDUN has been updated, and it is now available under three progressively deeper flavours: LINDDUN GO, LINDDUN PRO and LINDDUN MAESTRO. In particular, LINDDUN GO comes in the form of a card deck representing the most common privacy threats; LINDDUN PRO is more systematic and

exhaustive, supported by the knowledge base; LINDDUN MAESTRO targets an enriched system description to enable more precise threat elicitation, yet it is still under development.

The first notable difference with the old version lies in the acronym, which puts more emphasis on the privacy threat types rather than on the privacy properties affected by threats. In fact, for the sake of comparison, the acronym that was previously expanded as *Linkability*, *Identifiability*, *Non-repudiation*, *Detectability*, *Disclosure of information*, *Unawareness*, and *Non-compliance*, has now been revised as follows:

- *Linking* Associating data items or user actions to learn more about an individual or group.
- *Identifying* Learning the identity of an individual.
- *Non-repudiation* Being able to attribute a claim to an individual.
- *Detecting* Deducing the involvement of an individual through observation.
- *Data Disclosure* Excessively collecting, storing, processing or sharing personal data.
- *Unawareness & Unintervenability* Insufficiently informing, involving or empowering individuals in the processing of personal data.
- *Non-compliance* Deviating from security and data management best practices, standards and legislation.

The framework considers the state-of-the-art privacy threat types according to the privacy threat properties introduced by Pfitzmann [46]. These are categorised as hard privacy and soft privacy properties. In particular, unlinkability, anonymity and pseudonymity, plausible deniability, undetectability and unobservability, and confidentiality (hiding data content, including access control) are under the umbrella of hard privacy; user content awareness (including feedback for user privacy awareness, data update and expire) together with policy and consent compliance are, on the other hand, soft privacy properties.

LINDDUN provides a set of threats specific to privacy, named as "threat catalogue", in the form of threat trees. These privacy threat trees are inspired by the Security Development Lifecycle (SDL) [30] and reflect common attack patterns [64] on the basis of state-of-the-art privacy developments, structured according to LINDDUN or STRIDE threat category and, in the previous version of LINDDUN, also to Data Flow Diagram (DFD) element type. In fact, the consideration of the DFD interactions has become more implicit in the new version of the framework, as the threat trees have become independent from the DFD element type, thus resulting in a significant diminution of the number of nodes as a side effect. The new guidance on how to link the Data Flow Diagram interactions rests now solely on the LINDDUN mapping table.

Threat trees provide a formal way to describe the security of systems based on a variety of attacks. Basically, the root node represents the ultimate goal, e.g., the threatening to a property, the children nodes embody different ways of achieving that goal, i.e., refinements, hence leaves represent basic-level attacks that can not be further refined. In addition, non-leaf nodes can be conjunctive (logic AND) or disjunctive (logic OR) [54].

In the new version of LINDDUN, threat trees provide support to reason about applicability (criteria), factors that determine threat impact (impact), and examples of each characteristic pertaining to the threat (examples). The framework provides a different view of the threat trees in terms of detail, as it is possible to consult each tree at three different levels: Basic, Examples, All details.

An example tree is presented in Fig. 1 for the Linking threat, which can be achieved through $L.1$ "Linked data", e.g., IP address, and $L.2$ "Linkable data", e.g., browser fingerprint. Both of these provide various attack paths which are not necessarily limited to the LINDDUN property analysed, i.e., Linking could lead to Identifying threats if we consider $L.1.1$ "Unique identifier".

Our previous work [51] confirmed that the new version of LINDDUN represents a step forward from a GDPR perspective, as we can identify an increment of nodes in two LINDDUN privacy threat types, i.e., Unawareness & Unintervenability (threats against data subject rights) and Non-compliance (violations against data protection principles), which tightly align with the European regulation by including as many as 17 threats. In the previous version of LINDDUN, these two types were already bound to soft privacy, but only included 9 threats. Moreover, these soft privacy threats were lacking relevant aspects, such as those related to data subject controls, consent, and violation of regulations, which are now caught by the new threat knowledge base. On the other hand, the remaining types target more technical privacy threats, gathered under the umbrella of hard privacy, and as such contribute more directly to the selection of "appropriate technical and organisational protection measures".

Despite LINDDUN threat trees may lack some formal semantics and have minimal selection criteria to express potential threats [66], they still provide a valuable overview of potential threat types that seeks to be general, hence are suitable for a privacy threat analysis of any application domain. Moreover, the application of LINDDUN may lead to a high number of threats that may not be relevant, feasible, or important, thereby being labor-intensive and time-consuming, especially for complex or large systems [65]. Hence, the advantage of having a catalogue of privacy threats, which are broad and applicable to various domains, may result in the problem of threat explosion.

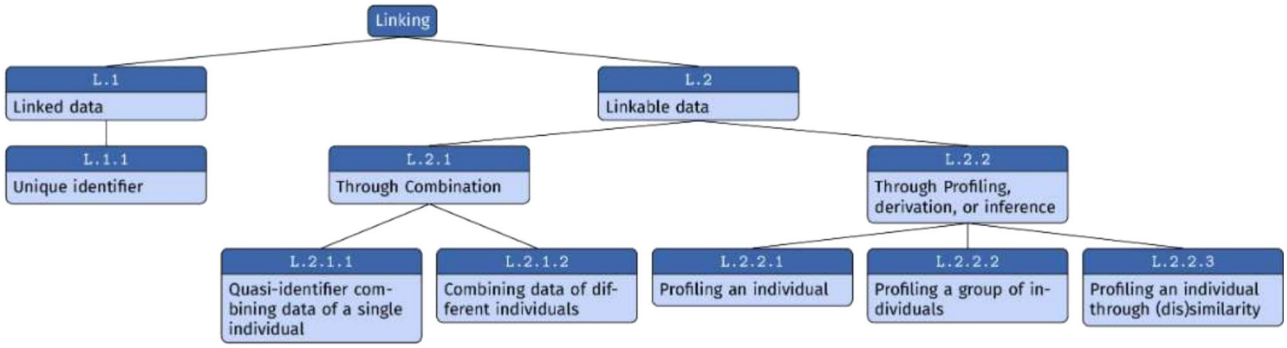

Fig. 1 Example of a LINDDUN threat tree: linking

4 The development and scientific reasoning of SPADA

This Section presents the scientific reasoning behind the development of the SPADA methodology. The development of SPADA is guided by principles from Design Science Research (DSR) [28, 45]. This approach emphasises the creation and iterative refinement of artefacts designed to solve identified problems. In the following, we shall leverage the steps inspired by DSR and specifically take them towards the development of SPADA, namely: Problem Identification and Motivation, Definition of Objectives, Design and Development, Demonstration, Evaluation, Communication.

4.1 Problem identification and motivation

We conducted a literature review to identify existing gaps in threat modelling studies and frameworks (§2), specifically their limitations in addressing both soft and hard privacy threats comprehensively. Existing methodologies such as LINDDUN and best-practices documents such as the reports from ENISA either lack of a focus on the target domain in its entirety (focus only on domain-independentness) or mainly consider cybersecurity aspects rather than privacy ones.

4.2 Definition of objectives

On the basis of the identified gaps, we set out to develop a methodology that is modular, comprehensive, and aligned with GDPR requirements. The primary objective is to create a flexible framework that can be adapted to various domains, such as smart cars and smart homes, yet at the same time providing a domain-independent threat knowledge base.

4.3 Design and development

As we shall detail below, the design of the SPADA methodology involves selecting the variable elements of threat modelling, defining the operations to handle multiple sources of documentation with attention to redundancy, and eliciting the steps for the proper execution of SPADA.

4.4 Demonstration

As we shall detail below, we apply the SPADA methodology to two timely and relevant domains, i.e., smart cars and smart homes, to demonstrate the practical applicability of SPADA. An initial partial validation of the demonstrations provides feedback for the practicality and tangibility of the results.

4.5 Evaluation

The SPADA methodology and its demonstration are evaluated to compare the results with prior work and analyse how SPADA changes the state of the art, as we shall detail below.

4.6 Communication

The SPADA methodology and the results of its demonstration are documented and communicated through the present manuscript and the repository available online [48].

5 The variable elements of threat modelling

This Section identifies the variable elements of threat modelling, i.e., the elements that contribute to model threats in general. The variable elements are the Source of documentation, the Property within privacy, the Application domain, the Detail (level of), and the Agent(s) raising the threats. Each of the variable elements is discussed below.

5.1 The source of documentation

In threat modelling, the knowledge base is crucial for both threats and assets to be elicited. Threats and assets may be derived from different sources, e.g., state-of-the-art reports, scientific contributions, guidelines, et cetera. There-

fore, the Source of documentation (or document source) of the threats/assets that the threat modelling seeks to gather can be either *internal* or *external* to the analyst's institution. In the case of internal document source, threats/assets may arise from the analyst's expertise, knowledge of the particular institutional context, or insights into the specific system or domain being assessed. On the other hand, the external document source involves gathering threats/assets from external references, such as established best practices or recognised industry standards. This allows the analyst to leverage existing knowledge and insights from a broader community of experts.

A mix of both internal and external document sources may also be possible, for example, when the analyst enucleates a new threat/asset being inspired from one or more external sources. In such a case, we refer to the document source of that threat/asset as *hybrid*. Furthermore, the document source variable provides the means to keep track of the version of the threats, for example, the *year* in which the specific threat list is published. Moreover, when considering two or more different document sources, it may likely happen that some threats within such lists are inherently embraceable. Hence, as we shall see below, the embrace operation remains crucial for leveraging various document sources.

5.2 The property

Privacy relates to the control that individuals have over their personal information, including how it is collected, used, and shared. According to the state of the art [13, 16], we can distinguish between two degrees of privacy, i.e., hard privacy and soft privacy. We identify in such properties the second variable element to build a privacy threat model, and contend that each of them deserves a specialised treatment.

Hard privacy refers to data minimisation, based on the assumption that personal data is not disclosed to third parties. The threat model includes service provider, data holder, and adversarial environment, where strategic adversaries with certain resources are motivated to breach privacy, similar to security systems [16]. Examples of hard privacy measures include anonymisation of data, data minimisation, and data retention policies, including the use of algorithms such as k-anonymisation, t-closeness and differential privacy. For example, a company that collects user data may anonymise the data before sharing it with third parties, ensuring that the users' identities remain protected. Similarly, a company may limit the amount of data it collects, processes, or stores to a minimum, and may have policies in place for deletion once processing those data are no longer necessary.

Soft privacy, on the contrary, is based on the assumption that the data subject lost control of their personal data and has to trust the honesty and competence of data controllers [16]. Examples of soft privacy measures include transparency and consent mechanisms, data subject access rights, and privacy impact assessments. For example, a company may provide clear and concise privacy notices to inform users about how their data will be used and shared. They may also obtain users' consent before using their data for purposes beyond the original scope. Additionally, companies may (and should, to be GDPR-compliant) provide users with the right to access, modify, or delete their personal data.

In summary, while hard privacy focuses on minimising the risks associated with the collection and retention of personal data, soft privacy focuses on the appropriate use and sharing of personal data while respecting individuals' rights to control their data. It is clear that, in addition to hard privacy and soft privacy, *cybersecurity* plays a major, complementary role in terms of protection against the unauthorised access of data.

As we shall see below, the SPADA methodology pays specific attention to both incarnations of privacy.

5.3 The application domain

The application domain in threat modelling identifies two prevailing approaches, i.e., the domain-dependent and domain-independent ones. Domain-dependent threat modelling is specific to a particular application domain, such as healthcare, finance, or automotive, and it takes into account the unique characteristics of the domain itself, thus it may be more accurate and effective. On the other hand, domain-independent threat modelling is not specific to the application domains and can be applied to a wide range of systems. It uses general threat categories, such as spoofing, tampering, and repudiation in STRIDE, to identify and prioritise threats.

A general threat knowledge base comes particularly useful in situations where there is no prior knowledge of the system or domain. LINDDUN, for example, currently takes a domain-independent approach. Domain-dependedness can be achieved by starting from a general threat knowledge base and by associating it to domain-specific characteristics, i.e., assets. This is what, for example, ENISA did in the studies reported in "Threat Landscape and Good Practice Guide for Smart Home and Converged Media" [17] and in "Good practices for security of smart cars" [19]. They addressed the domain-dependent vs domain-independent dilemma by using the combine operation that, as we shall see below, also SPADA adopts.

In consequence, the analyst has the possibility to follow two directions. The first direction implies the elicitation of domain-dependent threats starting from domain-independent threats. The other direction is the opposite, implying the elicitation of domain-independent threats starting (also) from domain-dependent threats. As we shall see below, SPADA is demonstrated on the first direction in the present manuscript.

5.4 The detail (level of)

Another variable element of threat modelling derives from the level of detail—of the statement describing a threat/asset. For example, "Unchanged default password" is certain to be more detailed than (the more abstract) threat "Human error", and clearly the former would be chosen should the analyst prefers a detailed level of granularity, whereas the latter would be selected if the focus were on a more abstract level. Normally, the analyst strives to choose a consistent level of detail till the end of the exercise. The level of detail becomes relevant in the context of threat modelling and, subsequently, in risk assessment exercises with respect to the likelihood estimation of a threat.

The concepts of *hyponym* and *hypernym* play an important role in understanding the level of detail in a statement, and refer to the "type of" semantic relation between terms [63]. The relationship is asymmetric, meaning that while a hypernym may include many hyponyms, i.e., a hypernym is an umbrella term, a hyponym with a clear semantics may only have one direct hypernym, which, in turn, is a hyponym for another hypernym, in a transitive relation. For example, Ferrari and Lamborghini are hyponyms of the automobile hypernym, itself a hyponym of the vehicle hypernym.

Analogously, *meronym* and *holonym* refer to the "part of" semantic relation, which is asymmetric and transitive too. For example, oven and dishwasher are meronyms of the kitchen holonym, itself a meronym of the house holonym.

In the case of threats, a *higher level of detail* demands the choice of a hyponym/meronym rather than a hypernym/holonym, possibly over several rounds, which implies that the analyst is able to estimate the likelihood of the given threat with more precision. However, an excessive level of detail leads to the degeneration of the threat to a "measurable event", hence to an exact assignment of the likelihood, that is either the bottom or the top in the given range. If the analyst's aim is to obtain a checklist of measurable events, a higher level of details represents the best option. Therefore, the most appropriate level of detail, i.e., the choice of employing hyponyms/meronyms or hypernyms/holonyms, should be considered within the main picture, and the analyst will choose it with some inevitable bias. As we shall see below, the SPADA methodology addresses this variable by supporting the definition of a target level of detail, which may, for example, be set to either *Abstract* or *Detailed*. This allows the analyst to choose the level of detail, thus they can adjust the threats/assets from the document sources that they considered towards that level of detail. The target level of detail can be tailored using the operations of embrace, rename, and discard defined by SPADA, as we shall see below. This process can be iterative, meaning the analyst explores semantic relationships between terms in multiple rounds, so as to produce a list of threats/assets with a consistent and coherent level of detail that aligns with their specific objectives. In fact, when considering multiple threats/assets, understanding the semantic relations between them helps determine the appropriate level of detail for analysis and response. By leveraging the semantic relations mentioned above, the analyst can focus more comprehensively on higher-level threat/asset categories or overarching threat scenarios. On the other hand, all this can also guide the analyst towards specific threat/asset instances or components, enabling a more targeted and detailed investigation.

5.5 The agent(s) raising the threats

The agent(s) raising the threats, or more commonly known as threat agents, are individuals, groups, or systems that have the capability to exploit vulnerabilities and cause harm to a system or organisation. In threat modelling, understanding the capabilities, motivations, and objectives of threat agents is crucial for identifying and prioritising threats. There are different types of threat agents, including insiders, outsiders, script kiddies, hacktivists, cybercriminals, and nation-state actors.

In the context of privacy threat modelling, we refer to a threat agent as any entity, individual or group, who poses a threat to an individual's privacy. Unlike the security literature, which traditionally refers to such entities as "adversaries" or "attackers", here the term threat agent is broader, i.e., not limited to malicious actors only. In fact, we also consider three additional actors directly from GDPR: data controller, data processor, and third party as threat agents. Therefore, a threat agent can be one or more of the following entities:

- *Attacker* Anyone, including an insider, or anything, including malware, acting with malicious intent to compromise a system to breach users' privacy.
- *Data controller* The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data;
- *Data processor* A natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;
- *Third party* A natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data.

This taxonomy allows us to better model the data subject's perspective during the threat modelling exercise. For example, it might be relevant to understand whether to classify a threat against data as a data breach or data leak. Article

4.12 of GDPR defines "*any breach of security leading to the accidental or unlawful destruction, loss, alteration or unauthorised disclosure of or access to personal data transmitted, stored or otherwise processed*", while ENISA specifies a data leak as an event that can cause the unintentional release of sensitive, confidential or protected data due to, for example, misconfigurations, vulnerabilities or human errors [20].

As we shall see below, the SPADA methodology assumes that one or a mix of multiple threat agents may be involved in the threat scenario.

6 The SPADA methodology

This Section advances the SPADA methodology for threat modelling. SPADA incorporates both domain-independent and domain-specific knowledge and considers the potential consequences on the security and privacy of individuals as its cornerstone. The SPADA methodology consists in the execution of methods of threat elicitation and asset collection. It rests on variables to define operations to take through precise steps. By leveraging the specific principles discussed above (§4), this Section details SPADA with its five variables, four operations and four steps.

6.1 The SPADA variables

The SPADA methodology adopts all the five variable elements, detailed in Sect. 5, that contribute to model (security and privacy) threats and names them *variables*. These variables compose the acronym of SPADA itself:

- *S*ource of documentation: the source(s) from which threats and assets may be derived, i.e., internal, external, or hybrid.
- *P*roperty: the specific version of the target property, i.e., soft privacy, hard privacy, or cybersecurity.
- *A*pplication domain: the domain dependentness, i.e., domain-indepedent or domain-dependent (e.g., smart car, smart home, etc.).
- *D*etail (level of): the style of the statement describing a threat/asset, which can be either abstract or detailed.
- *A*gent(s) raising the threats: typically including agents such as attacker, data controller, data processor, and/or third party, while a mix of them is also possible.

The inclusion of five essential variables in the SPADA methodology orients the analysis, thus providing practical guidance to the analyst.

6.2 The SPADA operations

SPADA adopts the following four operations:

- *Combine* It instantiates a domain-independent threat with the domain-specific assets that may be affected by such threat, typically by continuing the threat description with an explicit reference to the asset. This operation allows for the elicitation of both a general threat knowledge base and domain-dependent threats, which are derived from the first. An application of the combine operations is referred to as *combination*.
- *Embrace* It merges multiple threats into a single threat, typically by a new threat description that embodies the semantics of the given threats. The resulting threat description is determined by the input threat with the most pertinent level of detail, or it defaults to the description of the first threat if detail levels are similar. This operation can be used iteratively. An application of the embrace operation is hereby referred to as *embracing*.
- *Rename* It modifies the description of a threat. If the default description is considered incomplete or requires adjustment, the rename operation provides a way to refine the level of detail in the threat description. This results particularly useful when the analyst wants to further refine a threat description produced by the embrace operation.
- *Discard* It excludes a threat from the current analysis. The discarded threat may be kept in a reserve list for potential future review. This operation is essential when a threat is deemed irrelevant or inapplicable to the domain, e.g., it pertains strictly to a property that is not the target property or is not relevant to the specific target system.

6.3 The SPADA steps

Figure 2 depicts the SPADA methodology, and in particular its very steps are defined as follows:

Step 1 Variable setup
Step 2 Domain-independent threat elicitation
Step 3 Domain-dependent asset collection
Step 4 Domain-dependent threat elicitation

The methodology starts with Step 1, which consists in the choice of the values of the five variables. Step 2 involves the collection of domain-independent threats from relevant document sources. The variables that influence this step are document source, specific property, threat agents, and level of detail, whilst the operations that may be involved during this step are embrace, rename and discard. At this point, if the analyst is interested in eliciting threats that are specific

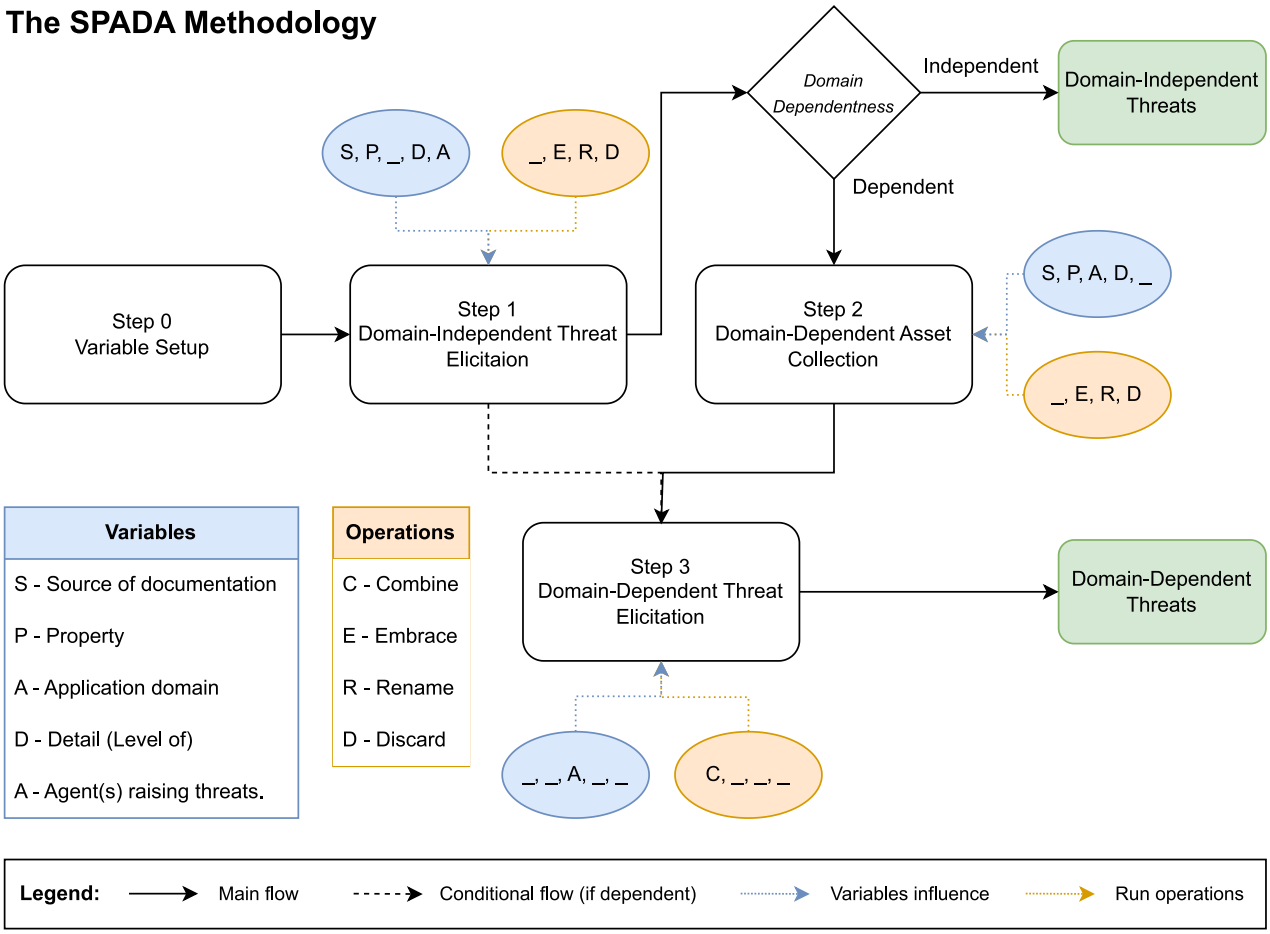

Fig. 2 Diagram of the SPADA methodology for threat modelling

to the application domain, i.e., the application domain variable is chosen to be domain-dependent, then the execution of the remaining steps continue. Step 3 consists of the collection of a list of assets for the target domain from relevant document sources. The variables that influence this step are document source, specific property, and threat agents. Also this step applies operations embrace, rename and discard. Finally, Step 4 aims at producing a list of domain-specific threats. In particular, for each domain-independent threat elicited in Step 2, this step associates to it the assets that were enumerated in Step 3. The sheer association expresses the object of the threat that was domain-independent in the first place, thereby making it domain-dependent. In other words, the domain-independent threat is instantiated *appropriately*, i.e., *over each of the assets it affects*, not necessarily all assets, ultimately producing a domain-dependent threat. The only variable that influences the execution Step 4 is the application domain, while the only operation that is needed is the combine one.

In summary, SPADA produces either a list of domain-independent threats, at Step 2, or adds a list of domain-dependent threats, at Step 4. While relevant examples will be given below, if $dit_1, \ldots, dit_n$ is the list of domain-independent threats produced by Step 2, then the number of domain-dependent threats that arise can be calculated as follows:

$$affected\_assets(dit_1) + \ldots + affected\_assets(dit_n).$$

7 Demonstration

This Section demonstrates the SPADA methodology according to the following strategy. We choose to apply SPADA to two timely and relevant application domains, i.e., smart car and smart home.

In particular, we select *external* document sources for both domains. For the sake of abbreviation, "ENISA TT" stands for the ENISA "Threat Taxonomy v2016" [18] report, "ENISA SC" for the "Good practices for security of smart cars" [19] report, "ENISA SH" for the "Threat Landscape and Good Practice Guide for Smart Home and Converged Media" [17] report, "OWASP" for the "Calculation of the

complete Privacy Risks list v2.0" [42] document, with the remaining labels "LINDDUN" [60], "Bella et al." [4] and "Kavallieratos et al." [33] being self-explanatory.

The document sources for the smart car domain are ENISA TT, ENISA SC, OWASP, Bella et al. and LINDDUN. The rationale behind these choices is that ENISA TT, OWASP, and LINDDUN offer domain-independent privacy threats, hence they can be leveraged for Step 2 in SPADA. Furthermore, Bella et al. as described in Sect. 2.2, proposed a double assessment, combining an asset-oriented ISO approach with a threat-oriented STRIDE approach. The asset-oriented ISO approach provided a categorisation of assets in the smart car domain, hence Bella et al. can be leveraged for the collection of the assets in Step 3. Moreover, ENISA SC provides a taxonomy of threats and assets for smart cars. Notably, ENISA adopts an approach similar to Step 4 in SPADA, i.e., the proposed domain-independent threats are associated by combination to the assets. For such reason, ENISA SC can be leveraged in Step 2 for the elicitation of domain-independent threats, yet also in Step 3 for the collection of the assets.

The document sources for the smart home domain are ENISA TT, ENISA SH, OWASP, Kavallieratos et al., and LINDDUN. While the same considerations for the smart car domain apply for ENISA TT, OWASP, and LINDDUN, the rationale behind the choice of ENISA SH is that it includes a taxonomy of the key assets in the smart home domain, useful for Step 3 in SPADA, together with Kavallieratos et al. that, as described in Sect. 2.3, examined existing dynamic risk assessment methodologies and provided a taxonomy of assets for smart homes.

In addition, for each of the two domains, we want to analyse both soft and hard privacy, hence the property must be flipped to consider soft privacy first and hard privacy secondly. As a result, a total of four exercises compose the whole demonstration.

As we identify the specific domains, i.e., smart car and smart home, it follows that the application domain is domain-dependent.

Also, we want to analyse threats with a lower level of detail to cover both the smart car and smart home domains at a macroscopic level.

In the pursuit of completeness, we consider all of the threat agents advanced in Sect. 5.5, i.e., attacker, data controller, data processor, and third party.

The full results of the demonstration are available online [48] and are conveniently structured as follows: an Excel file containing the list of threats extracted from the document sources (before applying the SPADA methodology); an Excel file containing the list of the elicited domain-independent threats for soft and hard privacy; an Excel file containing the list of the collected smart car assets along with a table presenting the list of the elicited domain-dependent threats for smart cars; an Excel file containing the list of the collected smart home assets along with a table presenting the list of the elicited domain-dependent threats for smart homes.

S	External *(ENISA TT, ENISA SC, OWASP, Bella et al., LINDDUN)*
P	Soft Privacy
A	Domain-Dependent *(Smart Car)*
D	Abstract
A	Attacker, Data Controller/Processor, Third Party

Fig. 3 Variable setup for smart car—soft privacy

7.1 Smart car—soft privacy

7.1.1 Variable setup

The first application for the smart car domain sets the variables discussed through Sect. 6 as depicted in Fig. 3.

7.1.2 Domain-independent threat elicitation

Soft privacy is the target property, therefore we must consider the LINDDUN threats that refer to such property, i.e., U(nawareness & unintervenability) and N(on-compliance), as a first *external* document source. For each node of the U-N property trees, we annotate the pertaining threat in a table. It is convenient to provide a brief and general explanation of these threats, referring to the new descriptions provided by their sources. In particular, U(nawareness & unintervenability) refer to situations where individuals are not adequately informed, involved, or empowered in the processing of their personal data. N(on-compliance) refers to situations where a system deviates from security and data management best practices, standards, and legislation. It primarily focuses on the organisational and operational management context in which a system or service operates.

Furthermore, we extend the list of domain-independent threats by adding other *external* document sources. In particular, our previous work [50] included the 8 threats that were found [49] to be outstanding with respect to the old version of LINDDUN. In detail, they account for the 2 threats from the ENISA SC report that fall under the "Legal" category, i.e., "Failure to meet contractual requirements" and "Violation of rules and regulations/Breach of legislation/Abuse of personal data", and the 6 threats from the OWASP document, i.e., "Consent-related issues", "Inability of user to access and modify data", "Insufficient data breach response", "Misleading content", "Secondary use", "Sharing, transfer or processing through 3rd party".

These threats relate to soft privacy as per the definition of soft privacy that we covered previously in Sect. 6. Moreover, some of them are embraceable with the new threat catalogue

proposed by LINDDUN. In particular, we notice that "Violation of rules and regulations/Breach of legislation/Abuse of personal data" is now *embraceable* with several threats such as "Regulatory non-compliance" and "GDPR"; "Consent-related issues" is now *embraceable* with "Invalid consent"; "Inability of user to access and modify data" with "Lack of data subject control"; "Insufficient data breach response" with "GDPR". Hence, we can discard those threats, since they are already contemplated in the new LINDDUN threat trees, and keep the following ones: "Failure to meet contractual requirements", "Misleading content", "Secondary use", "Sharing, transfer or processing through 3rd party".

Moreover, we also consider here the ENISA TT report as another *external* document source, as it is relevant to enrich the domain-independent threat knowledge base. We pick the threats that specifically target soft privacy. These can be found under the "Legal" category, i.e., "Violation of laws or regulations/Breach of legislation", "Failure to meet contractual requirements", "Unauthorized use of IPR protected resources", "Abuse of personal data", and "Judiciary decisions/court orders". Again, three of such threats are already included in the more recent ENISA SC report. In fact, "Failure to meet contractual requirements" is repeated and "Violation of laws or regulations/Breach of legislation" is embraced with "Abuse of personal data" into one single threat. Hence, we can add the following threats to the final list: "Unauthorized use of IPR protected resources", "Judiciary decisions/court orders". It is noteworthy that these additions are still possible without consequences on the domain variable, as such threats are general privacy threats, i.e., they ignore domain-specific entities. Hence, such threats can be analysed in relation with (virtually) any application domain.

For the sake of simplicity, we left the level of detail of the threats' descriptions unvaried. Arguably, the document sources that we considered share a level of detail that is comparable. In summary, we elicited a total of 23 soft privacy threats from the selected document sources, i.e., LINDDUN, ENISA (both ENISA TT and ENISA SC and OWASP. Table 1 shows such threats—the 6 that are highlighted are those that we do not deem embraceable with the current LINDDUN threats, hence represent our updated proposal for an extension to it. Moreover, while the 2 threats in italics are actually new (as they originate from the newly considered ENISA source, i.e., ENISA TT), the remaining 4 already were among the 8 that we suggested before [49]. It means that we managed to embrace half of the previous suggestions to current LINDDUN threats, something that we interpret as evidence that LINDDUN has been extended coherently with what we advocated.

Table 1 Domain-independent soft privacy threats elicited in Step 2

S	Domain-independent soft privacy threat
U	Unawareness of processing
	Unawareness as data subject
	Unawareness as a user sharing personal data
	Lack of data subject control
	Lack of data subject control–preferences
	Lack of data subject control–access
	Lack of data subject control–rectification/erasure
N	Regulatory non-compliance
	GDPR
	Insufficient data subject controls
	Violation of data minimization principle
	Unlawful processing of personal data
	Invalid consent
	Lawfulness problems not related to consent
	Violation of storage limitation principle
	Improper personal data management
	Insufficient cybersecurity risk management
ENISA	Failure to meet contractual requirements
	Unauthorized use of IPR protected resources
	Judiciary decisions/court orders
OWASP	Misleading content
	Secondary use
	Sharing, transfer or processing through 3rd party

7.1.3 Domain-dependent asset collection

For Step 3, we leverage two *external* document sources from the state of the art, i.e., the assets identified in the work proposed by Bella et al. [4] and ENISA's taxonomy of the key assets in the smart car domain included in ENISA SC. Bella et al. present the following list of assets:

- *Personally identifiable information* Any data that could potentially be used to identify a particular individual, such as full name, date, and place of birth, driving licence number, phone number, mailing, and email address.
- *Special categories of personal data* Data about the driver, e.g., racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, data concerning health or data concerning sex life or sexual orientation (Article 9 of GDPR).
- *Driver's behaviour* Driver's driving style, e.g, the way the driver accelerates, speeds up, turns, brakes.
- *User preferences* Data regarding cabin preferences, e.g., seating, music, windows, heating, ventilation and air conditioning (HVAC).
- *Purchase information* Driver's financial information, such as credit card numbers and bank accounts.

- *Smartphone data* Data that the vehicle and driver's smartphone exchange with each other via the mobile application and short-range wireless connections such as Wi-Fi and Bluetooth (contact book, phone calls, text messages).
- *GPS data* Vehicle geolocation history and route tracking.
- *Vehicle information* Vehicle information such as car-maker, model, vehicle identification number (VIN), licence plate and registration.
- *Vehicle maintenance data* Data about the maintenance and status of vehicle components such as kilometres travelled, tyre pressure, oil life, brake, suspension, and engine status.
- *Vehicle sensor data* Data analysed and calculated by car sensors, such as distance sensors, crash sensors, biometric sensors, temperature sensors and internal and external cameras.

The ENISA SC report focuses on Automated Driving System-Dedicated Vehicle (ADS-DS) [53], i.e., semi-autonomous and autonomous cars, and V2X communications, pertaining to SAE Level 4 and Level 5. The focus of the study is on smart cars that, as connected systems, have the necessary capabilities to autonomously perform all driving functions under certain (or all) conditions, and are able to communicate with their surroundings including other vehicles, pedestrians and Road-Side Units (RSU). Moreover, the key concepts analysed by ENISA do not only concern passenger cars but also commercial vehicles (e.g. buses, coaches, etc.), including self-driving, ride-sharing vehicles that can be shared with other users.

The assets proposed by ENISA are categorised in: "Car sensors and actuators", "Decision Making Algorithms", "Vehicle Functions", "Software management", "Inside vehicle Communication Components", "Communication Networks and Protocols", "Nearby External Components", "Network and Domain Isolation Features", Servers", "Systems and Cloud Computing", "Information", "Humans", "Mobile Devices". For the sake of brevity, we only quote the descriptions of the assets under the "Information" category:

- *Sensors data* Refers to data that is gathered by the different smart car sensors and which will be transmitted to the appropriate ECU for processing.
- *Keys and certificates* Refers to the different keys and certificates used for security purposes (such as authentication, securing the exchanges, secure boot, etc.). Keys are stored in devices embedded in the vehicle (e.g. ECU) and/or in servers depending on their use.
- *Map data* Refers to the information about the car environment. Map data allows us to increase the passenger safety by correlating its information with the sensor data. Contrary to GNSS, which gives only information about the geolocalisation, map data gives information about the surrounding environment.
- *V2X information* Refers to the various information exchanged via V2X communications (e.g. emergency vehicle approaching, roadworks/collision warning and traffic information).
- *Device information* Refers to the various information related to a device embedded in a smart car (e.g. ECU, TCU) or connected devices (e.g. smartphones, tablet). This includes information such as type, configuration, firmware version, status, etc., of different smart car sensors and which will be transmitted to the appropriate ECU for processing.
- *User information* Refers to smart cars user (e.g. driver, passenger, etc.) information such as name, role, privileges and permissions.

During the execution of this step, within the list proposed by Bella et al., we identified some assets that are embraceable with the ENISA taxonomy. In particular, "Personally Identifiable Information" is *embraceable* with "User information"; "Smartphone data" with "Device information"; "GPS data" with "Map data"; "Vehicle sensor data" with "Sensor data". Thereby, we explicitly picked the following assets from the article contribution: "Special categories of personal data", "Driver's behaviour", "User preferences", "Purchase information", "Vehicle information', "Vehicle maintenance data". The last two were not available before. The remaining assets, according to our scrutiny, are already contemplated in the ENISA taxonomy.

For the sake of simplicity, also in this case, we left the level of detail of the assets' descriptions unvaried, as the document sources that we considered share a level of detail that is comparable. Overall, we elicited a total of 43 assets, a small increase on the 41 that we had before [50].

7.1.4 Domain-dependent threat elicitation

In the last step, we conjugate the findings from the previous steps. For each domain-independent threat elicited in Step 2, we assign the assets from Step 3 that we deem to be potentially affected by that particular threat. In general, a threat may apply to multiple assets, therefore for some threat-asset pairs we annotate multiple assets or, in case all assets are affected, we add the description "All assets" for the sake of brevity. In particular, most assets that we deem to be potentially affected by the soft privacy threats fall under the ENISA category "Information".

While the full results are available online [48], we present an exemplification of some noteworthy domain-dependent

threats, with the additional aim of providing the rationale behind the related threat-asset(s) associations:

$dit\_sc\_s_{i_1}$ —*Unawareness of processing* refers to the lack of awareness or understanding about how personal data is being processed. It affects various assets, such as sensors data, map data, V2X information, device information, user information, special categories of personal data, user preferences, purchase information, vehicle information, and vehicle maintenance data.

$dit\_sc\_s_{i_2}$ —*Lack of data subject control—preferences* specifically refers to the lack of control individuals have over their preferences. It affects assets such as user preferences and purchase information. When individuals cannot control or manage their preferences effectively, their privacy in relation to their preferences can be at risk.

$dit\_sc\_s_{i_3}$ —*Regulatory non-compliance* encompasses all assets. It refers to the failure to comply with relevant privacy regulations or laws. When organisations do not adhere to the required privacy standards, all assets can be affected, leading to potential privacy breaches.

$dit\_sc\_s_{i_4}$ —*GDPR* is also associated with all assets. It specifically refers to non-compliance with the General Data Protection Regulation (GDPR), a data protection law in the European Union. Violations of GDPR can lead to severe penalties and legal consequences.

$dit\_sc\_s_{i_5}$ —*Violation of data minimization principle* refers to the violation of collecting and processing only the necessary data. It affects assets such as sensors data, map data, V2X information, device information, user information, special categories of personal data, user preferences, and purchase information, vehicle information, and vehicle maintenance data.

$dit\_sc\_s_{i_6}$ —*Unlawful processing of personal data* covers all assets. It occurs when personal data is processed unlawfully or without a legal basis. When personal data is processed in violation of applicable laws or regulations, it poses a significant privacy risk to all assets involved.

$dit\_sc\_s_{i_7}$ —*Lawfulness problems not related to consent* is associated with all assets. It highlights issues of lawfulness in data processing that are not specifically related to consent. These problems may include processing of personal data without a valid legal basis or exceeding the scope of permitted processing activities, such as automated decision-making on sensitive personal data.

S	External *(ENISA TT, ENISA SC, OWASP, Bella et al., LINDDUN)*
P	Hard Privacy
A	Domain-Dependent *(Smart Car)*
D	Abstract
A	Attacker, Data Controller/Processor, Third Party

Fig. 4 Variable setup for smart car—hard privacy

$dit\_sc\_s_{i_8}$ —*Improper personal data management* is associated with user information and special categories of personal data. It signifies improper management practices regarding personal data, including inadequate safeguards, inappropriate handling, or unauthorised access. Improper data management can lead to privacy breaches, data leaks, or unauthorised use of sensitive information.

$dit\_sc\_s_{i_9}$ —*Failure to meet contractual requirements* refers to a breach of contractual requirements by Tier 1 and/or Tier 2 car components or software suppliers, thus encompassing all assets. Such threat may lead to financial, safety, privacy and/or operational impacts.

$dit\_sc\_s_{i_{10}}$ —*Sharing, transfer or processing through 3rd party* refers to the sharing or transferring of various assets to third parties that increases the likelihood of unauthorised access, misuse, or breaches. It is clear that the affected assets belong to the ENISA "Information" category and include special categories of personal data, driver's behaviour, user preferences, purchase information, vehicle information, and vehicle maintenance data.

As an outcome of this exemplification, the resulting number of domain-dependent threats would be:

$$affected\_assets(dit\_sc_{s_1}) + \ldots +$$
$$+ affected\_assets(dit\_sc_{s_{10}}) = 10 + 2 + 43 +$$
$$+ 43 + 10 + 43 + 43 + 2 + 43 + 12 = 251$$

7.2 Smart car—hard privacy

7.2.1 Variable setup

The second application for the smart car domain sets the variables discussed through Sect. 6 as depicted in Fig. 4.

7.2.2 Domain-independent threat elicitation

Hard privacy is the target property, therefore we must consider the LINDDUN threats that refer to it, i.e., L(inking),

I(dentifying), N(on-repudiation), and D(etecting). For each node of the L-I-N-D property trees, we annotate the pertaining threat in a table. It is convenient to provide a brief and generic explanation of these threats, referring to the new descriptions provided by their sources.

In particular, L(inking) refers to the process of associating data items or user actions in order to gain a better understanding of an individual or a group. I(dentifying) refers to the process of learning the identity of an individual. While many systems require the identification of data subjects, identifying threats emerge when the identity can be unintentionally or undesirably revealed through leaks, deduction, or inference. N(on-repudiation) refers to the ability to attribute a claim, such as an action, statement, or event, to an individual. This eliminates the possibility of plausible deniability, leaving individuals exposed to potential consequences, such as legal prosecution, particularly in cases involving whistleblowers. D(etecting) involves deducing the involvement of an individual through observation. Unlike other threats, detecting does not necessarily require access to the actual data itself. Simply knowing that the data exists is often sufficient to infer additional sensitive information.

Furthermore, we extend the list of domain-independent threats by adding other *external* document sources. In particular, we leveraged again the OWASP document as an additional document source. We identified two threats, i.e., "Insufficient data quality" and "Data aggregation and profiling", that refer to hard privacy as per the definition of hard privacy covered previously in Sect. 6. Moreover, "Data aggregation and profiling" is embraceable with the new threat catalogue proposed by LINDDUN, in particular with "Linkable data—Through profiling, derivation, or inference". Hence, we only added "Insufficient data quality" to the domain-independent hard privacy threat list. Again, these additions are still possible without consequences on the domain variable, as such threats are general privacy threats, hence they can be analysed in relation with (virtually) any application domain.

For the sake of simplicity, we left the level of detail of the threats' descriptions unvaried again, since the document sources that we considered share a level of detail that is comparable. In summary, we elicited a total of 29 hard privacy threats from the selected document sources, i.e., LINDDUN and OWASP. Table 2 shows such threats.

Table 2 Domain-independent hard privacy threats elicited in Step 2

S	Domain-independent hard privacy threat
	Linked data
	Linked data—unique identifier
	Linkable data
	Linkable data—through combination
	Quasi-identifier combining data of a single individual
	Combining data of different individuals
	Linkable data–through profiling, derivation, or inference
	Profiling an individual
	Profiling a group of individuals
L	Profiling an individual through (dis)similarity
	Identified information
	Processing of identified information
	Identified information in metadata
	Identifiable information
	Pseudonym
	Pseudonym—identifier
	Pseudonym—quasi-identifier
	Revealing attributes
I	The data subject is distinguishable from others
	Attributable data evidence
	Attributable data evidence—data
	Attributable data evidence—signed data
	Attributable data evidence—metadata
	Attributable data evidence—embedded/Hidden data
N	Attributable action side-effect evidence
	Observed communications
	Application side-effect
D	System responses
OWASP	Insufficient data quality

7.2.3 Domain-dependent asset collection

For the sake of brevity, we omit repetition with the previous exercise. In particular, Step 3 considers the same domain-dependent assets that we already discussed in Sect. 7.1.

7.2.4 Domain-dependent threat elicitation

In the last step, we conjugate the findings from the previous steps. For each domain-independent threat elicited in Step 2, we assign the assets from Step 3 that we deem to be potentially affected by that particular threat.

While the full results are available online [48], for the sake of brevity we leave the exemplification of some noteworthy domain-dependent threats in Appendix A.

7.3 Smart home—soft privacy

7.3.1 Variable setup

The first application for the smart home domain sets the variables discussed through Sect. 6 as depicted in Fig. 5.

S	External *(ENISA SH, ENISA TT, Kavallieratos et al., OWASP, LINDDUN)*
P	Soft Privacy
A	Domain-Dependent *(Smart Home)*
D	Abstract
A	Attacker, Data Controller/Processor, Third Party

Fig. 5 Variable setup for smart home—soft privacy

7.3.2 Domain-independent threat elicitation

For the sake of brevity, we omit repetition with the previous exercise. In particular, Step 2 considers the same domain-independent threats that we already discussed in Sect. 7.1.

7.3.3 Domain-dependent asset collection

For Step 3, we leverage two *external* document sources from the state of the art, i.e., the assets identified in the work proposed by Kavallieratos et al [34] and ENISA's taxonomy of the key assets in the smart home domain included in ENISA SH. The former presents the following list of assets:

- *User credential* Credentials that grant access to the smart home system and its functionalities. These may include usernames, passwords, and authentication tokens.
- *Information collected by smart devices* Data gathered from various sensors and devices within the smart home, providing insights into environmental conditions, usage patterns, and user behaviour.
- *Smart home status information* Real-time updates about the operational status of the smart home's systems, devices, and components, helping users monitor and control their home remotely.
- *Information about the installed assets* Details about the types, models, and configurations of devices and systems integrated into the smart home ecosystem.
- *Logs information* Records of activities, events, and interactions within the smart home environment, aiding in troubleshooting, analysis, and security monitoring.
- *Location tracking information* Data that tracks the physical location of occupants and objects within the smart home, enabling personalised services and context-aware automation.
- *Video, picture, voice information* Multimedia content captured by cameras, microphones, and other sensory devices, serving purposes such as surveillance, communication, and entertainment.
- *Health information* Sensitive data related to occupants' health, well-being, and medical conditions, often used to enable personalised healthcare services.
- *Billing data* Information related to payment transactions and usage charges for smart home services, including utility consumption and subscription fees.
- *Profile data* Personalised user profiles that store preferences, settings, and usage history to tailor the smart home experience to individual needs.
- *IoT smart devices* Devices equipped with sensors, actuators, and communication capabilities, contributing to the smart home's interconnected ecosystem.
- *IoT hubs* Centralised devices that facilitate communication and coordination among multiple IoT devices within the smart home network.
- *IoT gateways* Devices that connect the smart home network to external networks, such as the Internet, enabling remote access and control.
- *Sensors/actuators* Devices that sense and measure physical properties (sensors) or perform actions based on received instructions (actuators), contributing to automation and control.
- *Cloud server* Remote servers hosted in the cloud that store data, provide processing power, and enable remote access to and management of smart home services and devices.

The ENISA SH report focuses on smart homes in a broad manner, as smart homes are equipped with interconnected sensors, systems, and devices to provide automation, monitoring, and control through various means, such as computers, smartphones, and the internet. The report highlights three different models for smart homes, yet common installation of smart homes is likely to involve a blend of these three models.:

- *Fully decentralised model* In this model, each smart device operates autonomously and connects to the internet via the home network. Data transmission occurs through secure channels. Services may interact through communication between different providers or direct peer-to-peer integration. Security and privacy are not guaranteed by a single manufacturer, but rather rely on the overall network's considerations.
- *Local connectivity model* This option envisions smart devices connecting locally without relying on cloud services or a central gateway. Devices self-identify and form a solution by recognising each other's capabilities. However, this model faces technological barriers, a lack of shared protocols, and challenges in incorporating devices designed for internet connectivity.
- *Central hub or gateway model* A central hub or gateway coordinates various devices, integrating their services to offer advanced functionalities and value. This model, based on a central software system, can ensure secu-

rity and privacy since data remains confined to the home environment. Multiple devices, including smart TVs, smartphones, tablets, and wearables, can contribute to controlling the smart home.

The assets proposed by ENISA are categorised in: "Sensors", "Software", "Human–machine interface devices", "Home networking", "Audio/Visual", "Information Storage", "Home appliances", "Integrated Home services", "Robotics", "Tags and markers", "Building security", "Connected transportation", "Medical", "Information", "Management/operation", and "People/living".

From the latter category, we discarded "Pets" and "Plants" under the lines of GDPR appliance to only living individuals and, in terms of privacy, the inclusion of these two assets may require additional ethical reasoning that is out of scope in our study.

During the execution of this step, within the list proposed by Kavallieratos et al., we identified some assets that are embraceable with the ENISA taxonomy. In particular, "User credential" is *embraceable* with "Passwords" and "Access and payment credential for external accounts"; "Information collected by smart devices" with "Smart home setup/structure/inventory information"; "Smart home status information" is repeated; "Information about the installed assets" is *embraceable* with "Smart home setup/structure/inventory information"; "Video, picture, voice information" with the assets under the "Resources" category; "Health information" with "Medical"; "Billing data" is repeated as "Billing"; "IoT smart devices" is *embraceable* with "Home appliances"; "IoT hubs" with "Human–machine interface devices"; "IoT gateways" with "Interface to home gateway" and "Gateway"; "Sensors/actuators" with "Sensors" and "Actuators/Motors".

Also in this case, for the sake of simplicity, we left the level of detail of the assets' descriptions unvaried. Overall, we elicited a total of 127 assets.

7.3.4 Domain-dependent threat elicitation

In the last step, we conjugate the findings from the previous steps. For each domain-independent threat elicited in Step 2, we assign the assets from Step 3 that we deem to be potentially affected by that particular threat. In general, a threat may apply to multiple assets, therefore for some threat-asset pairs we annotate multiple assets or, in case all assets are affected, we add the description "All assets" for the sake of brevity.

Also in this case, while the full results are available online [48], for the sake of brevity we leave the exemplification of some noteworthy domain-dependent threats in Appendix A.

S	External *(ENISA SH, ENISA TT, Kavallieratos et al., OWASP, LINDDUN)*
P	Hard Privacy
A	Domain-Dependent *(Smart Home)*
D	Abstract
A	Attacker, Data Controller/Processor, Third Party

Fig. 6 Variable setup for smart home—hard privacy

7.4 Smart home—hard privacy

7.4.1 Variable setup

The second application for the smart home domain sets the variables discussed through Sect. 6 as depicted in Fig. 6.

7.4.2 Domain-independent threat elicitation

For the sake of brevity, we omit repetition with the previous exercise. In particular, Step 3 considers the same domain-independent threats that we already discussed in Sect. 7.1

7.4.3 Domain-dependent asset collection

For the sake of brevity, we omit repetition with the previous exercise. In particular, Step 3 considers the same domain-dependent assets that we already discussed in Sect. 7.3.

7.4.4 Domain-dependent threat elicitation

In the last step, we conjugate the findings from the previous steps. For each domain-independent threat elicited in Step 2, we assign the assets from Step 3 that we deem to be potentially affected by that particular threat.

Also in this case, while the full results are available online [48], for the sake of brevity we leave the exemplification of some noteworthy domain-dependent threats in Appendix A.

8 Partial validation

This Section presents a partial validation of the demonstration described above. The partial validation seeks to understand the practicality and relevance of the resulting threats by relying on the latest breaking news and articles about privacy incidents in the target domain. In particular, we employ classical web searches as a source of relevant information by building queries as "privacy smart car", "smart car breach", "smart car privacy", et similia for the smart car domain, and queries as "privacy smart home", "smart home breach", "smart home IoT privacy", et similia for the smart home domain, in the News search filter offered by

Google. The partial validation is conveniently structured in subsections reflecting each of the four applications from the demonstration above. Furthermore, we present some illustrative examples of news that matched with one or more of the proposed threats resulting from each exercise. The following examples extend our previous partial validation [50] and provide a different reading of the pieces of news in common, in light of the new threat list. For the sake of brevity, only the partial validation of the first exercise is given below, with the remaining left in Appendix B.

8.1 Partial validation of smart car—soft privacy

A data breach at Toyota Motor's Indian business [52] might have exposed some customers' personal information. "Toyota Kirloskar Motor (TKM) has been notified by one of its service providers of an incident that might have exposed personal information of some of TKM's customers on the internet". This perfectly embodies a threat that we find in Table 1, i.e., "GDPR", stemming from an inadequate response to a data breach that does not comply with GDPR.

Furthermore, we find another news that represents multiple threats: "GDPR", "Lack of data subject control", "Insufficient data subject controls" and "Violation of data minimization principle". The Dutch Data Protection Authority (DPA) investigated Tesla's camera-based "Sentry Mode" security system [21], which is designed to protect the vehicle against theft or vandalism while it is parked. It does this by taking footage with four cameras on the outside of the vehicle. This specific threat has now received a mitigation measure from the manufacturer, as the company altered security cameras to be more privacy-friendly and avoid GDPR violations. Originally, when Sentry Mode was enabled, this system was on by default. According to the news, the cameras continuously filmed everything around a parked Tesla and stored the last hour of footage each time.

In addition, we also found a review [59] that perfectly matched with the implications related to several soft privacy threats from the previous exercise. The article discusses a suggestion for a new feature to be added to the Ring Car Cam. The author proposes an Alexa-based voice command that would temporarily turn off the interior camera and microphone. This suggestion is based on the author's wife's volunteer work, which involves discussing private and privileged information about children's legal cases on the phone. The author's wife currently uses the physical privacy shutter to prevent the camera from recording video and audio inside the car. However, she sometimes forgets to flip the shutter up or down. Therefore, the author proposes a hands-free privacy trigger that would allow the user to enable or disable privacy mode with a voice command. This feature would eliminate the need for the user to physically manipulate the shutter, making it easier to maintain privacy while driving.

Moreover, we found a match for the "Improper personal data management" threat, as Toyota Japan [56] disclosed a significant data breach that occurred due to a cloud misconfiguration, resulting in the exposure of millions of customers' vehicle details over a decade. The exposed data included personal information, vehicle details, and videos.

Another discovery [12], related at least to the "Insufficient cybersecurity risk management" threat, revealed that BMW may have potentially exposed sensitive files and client data, including customer information, as a result of an unprotected environment and the exposure of configuration files on the official BMW Italy website. Although the information alone may not compromise the website, it could be used for reconnaissance purposes by attackers. As a typical example of interconnection between privacy and security, the exposed configuration file could have allowed threat actors to find other vulnerabilities and access the site's source code.

The same interconnection between privacy and security is also tangible in the National Highway Traffic Safety Administration (NHTSA) warning [35] to carmakers in Massachusetts not to comply with a state law that requires them to share more vehicular telematics data with third parties. This naturally embodies the "Judiciary decisions/court orders" threat. The NHTSA argues that the state law is pre-empted by federal law and could potentially allow attackers to remotely access and control cars, leading to safety risks. The law, known as the "right to repair" law, has been the subject of a court battle between carmakers and the state. The NHTSA's letter represents the federal government's direct involvement in the case and raises concerns about the potential dangers of open access to vehicle telematics. The litigation is likely to face further delays due to the NHTSA's intervention.

9 Evaluation

In this Section, we evaluate the main findings from the previous experiments, whose full versions are available online [48]. In summary, as a first key result, we produced a novel, refined list of soft privacy threats that are domain-dependent. In fact, we associated the general threat knowledge base pertaining to soft privacy, collected at the end of Step 2, with the smart-car-/smart-home-specific assets collected at the end of Step 3, thus obtaining domain-specific soft privacy threats for smart-car/smart-home devices with a homogeneous level of detail and dependent on the smart-car/smart-home domain, at the end of Step 4. For the sake of comparison between the two application domains, Table 3 summarises the results in terms of number of assets and domain-dependent threats elicited. As we may notice, the numbers of domain-dependent soft privacy threats for both smart cars and smart homes are significantly greater than hard privacy threats. Soft privacy encompasses legal and compli-

ance aspects, thereby threats related to this specific property target most of the assets and, as a consequence, the combinations produce a larger amount of domain-dependent threats.

It is important to emphasise that a crucial difference between the new list and the old list of threats was found: among the 8 threats added to the list in our previous work, 4 were deemed to be embraceable with the new LINDDUN threat catalogue. Hence, LINDDUN is clearly moving towards the direction that we hoped for, and we are confident that their threat knowledge base will continuously improve in such a positive direction. Also, this supports the case that embracing is relevant and useful, especially when the analyst considers different document sources. Table 4 presents a comparative analysis between the prior work [51] and the present work in terms of the number of domain-independent soft privacy threats elicited and the number of possible candidates for a LINDDUN extension for soft privacy. In detail, as stressed in Sect. 1.3, the 17 soft privacy threats that we made available when we adopted the previous version of LINDDUN are now extended to a total of 23. In consequence, because LINDDUN's soft privacy threats have increased from 9 to 17 over its two versions, as described in Sect. 3, our proposed extensions of LINDDUN have decreased from 8 to 6. This can be taken as an indication that LINDDUN has evolved in the direction we advocated.

As a second key result, hard privacy was treated by producing a list of hard privacy threats that are domain-independent. Similarly to our soft privacy threats list, we associated the general threat knowledge base pertaining to hard privacy (Step 2), with the smart-car/smart-home-specific assets collected (Step 3), thus obtaining domain-specific soft privacy threats for modern cars/smart home devices with a homogeneous level of detail and dependent on the smart-car/smart home domain (Step 4).

A confirmation of the practicality and relevance of both these soft and hard privacy threats for both the smart car and smart home domains was proven by means of web searches. This answers the research questions.

Our new lists of threats enriches the broader threat knowledge base in the smart car domain over both soft and hard privacy. Also, the smart home domain finally obtains a threat knowledge base over (soft and hard) privacy, thus filling the gap with the state-of-the-art threats that were merely security-oriented. In detail, Table 5 illustrates the document sources that compose our domain-independent threat knowledge base.

We recall that "ENISA TT" stands for the ENISA "Threat Taxonomy v2016" [18] report, "ENISA SC" for the "Good practices for security of smart cars" [19] report, "ENISA SH" for the "Threat Landscape and Good Practice Guide for Smart Home and Converged Media" [17] report, "OWASP" for the "Calculation of the complete Privacy Risks list v2.0" [42] document, with the remaining labels being self-explanatory. Notably, the ENISA TT report, OWASP and LINDDUN are in common for both the application domains as a further demonstration of their domain-independent nature, whilst the other document sources strictly refer to the specific application domain.

While we cannot claim that no more valid candidates exist, our final list of threats is complete with respect to the state-of-the-art knowledge base on soft and hard privacy threats. Our output is now available for the international community's evaluation.

9.1 How SPADA changes the SOTA

Prior to SPADA, existing methodologies such as LINDDUN and studies such as ENISA best practices and various academic contributions offered foundational approaches to threat modelling. However, these methodologies and studies have notable limitations. For example, while robust in certain areas and the recent focus on supporting domain-specific refinements, LINDDUN mostly remains domain-independent and often lacks the granularity needed for specific application domains (§3). Similarly, ENISA best practices and academic studies provide valuable insights but are not exhaustive in covering privacy threats, as they often focus more on cybersecurity aspects rather than comprehensive privacy concerns (§2).

SPADA addresses these limitations through guiding principles, i.e., its variables, thus ensuring a comprehensive modelling of security and privacy threats that is also adaptable to specific contexts. SPADA reduces the reliance on arbitrary decisions, thus enhancing the objectivity and reliability of the threat models. This adaptability marks a significant advancement in the field of security and privacy threat modelling, as it contributes to the elicitation of both domain-independent and domain-dependent threats, with a concrete support in the decision-making process. By collecting assets and combining them with domain-independent threats and the options for refiniment, SPADA provides a highly specific threat model for each target domain.

While other existing methodologies could, in principle, leverage the same variables as SPADA, it is a distinctive contribution of SPADA to provide clear guidelines for applying the five variables to tailor the threat elicitation process as the analyst demands. Table 6 provides a comparative analysis between SPADA and state-of-the-art methodologies referred from OWASP Threat Modelling Cheat Sheet [43], i.e., STRIDE, LINDDUN, OCTAVE, PASTA, and VAST. The comparison is based on the attention given to five key variables. Specifically, the symbol ✓ indicates full support for the variable, as it is a fundamental or inherent part of the methodology; △ represents partial support for the variable, either introduced as an add-on in later versions or lightly

Table 3 Assets and domain-dependent threats elicited over considered application domains

Domain	# Assets	# Soft privacy dd_t	# Hard privacy dd_t	# Total dd_t
Smart car	43	525	260	785
Smart home	127	1158	344	1502

Table 4 Comparisons of soft privacy threats and candidates for LINDDUN extension between prior and present work

	# Soft privacy threats	# Candidates for LINDDUN extension
Prior work	17	8
Present work	23	6

Table 5 Document sources distribution over considered application domains

Domain smart car smart home	Document source ENISA SH	ENISA TT	ENISA SC	Kavallieratos	OWASP	Bella	LINDDUN
Year	2014	2016	2019		2021	2023	

Table 6 Comparative Analysis of SPADA with SOTA methodologies based on the five variables

Methodology	S	P	A	D	A
SPADA	✓	✓	✓	✓	✓
STRIDE	×	△	×	△	△
LINDDUN	×	△	△	△	×
OCTAVE	△	△	△	×	×
PASTA	×	×	△	×	△
VAST	×	×	×	×	△

touched upon in the methodology; × denotes no explicit support for the variable.

SPADA is the first methodology to fully leverage the Source of documentation variable to gather, combine, and refine threats and assets. While OCTAVE's mere account for "detailed worksheets and questionnaires" indicates consideration of information sources, SPADA emphasises the general role of sources, which are not limited to working documents, but extend to all relevant existing documentation. The other methodologies do not achieve this.

Both STRIDE and LINDDUN define threat types (i.e., spoofing, tampering, etc. and linking, detecting, etc. respectively), yet they respectively focus on security and (both hard and soft) privacy. As such, they partially support the Property variable. OCTAVE, with its emphasis on security properties to protect critical assets, offers partial support for this variable, while PASTA and VAST, on the other hand, do not explicitly address the Property variable.

Both STRIDE and LINDDUN lack the capacity to tailor the threat modeling process to a specific domain, thereby they do not fully leverage the Application domain variable, as they remain domain-independent. Although LINDDUN was initially conceived as a domain-independent methodology, it has since evolved to support domain-specific refinements, thus offering partial support for the Application domain variable. OCTAVE, by focusing on organizational contexts and critical assets, and PASTA, by offering a stage-driven approach that integrates business objectives with technical details, partially support this variable. VAST, however, does not explicitly guide the adaptation of the methodology to specific application domains.

The Detail level for the description of threats and assets is explicitly guided only in SPADA, whereas STRIDE and LINDDUN do not consider semantic detail explicitly, as they only offer partial differentiation of detail levels indirectly within their threat catalogues (§3). OCTAVE, PASTA and VAST lack explicit mechanisms for addressing semantic detail in the description of threats and assets, leaving this variable unsupported.

While STRIDE allows some flexibility to consider threat agents during the process, LINDDUN does not explicitly address Agent(s) raising the threats as part of the threat elicitation process. OCTAVE, while identifying vulnerabilities and evaluating attack likelihood, does not consider specific adversaries. Conversely, PASTA and VAST adopt attacker-centric approaches, thus providing partial support for the Agent(s) raising the threats variable, without providing a fixed set of options as SPADA does.

10 Conclusions

This article faced the challenge of modelling privacy threats and addressed it by advancing the SPADA methodology for threat modelling. SPADA is general for security and privacy, and revolves around five variables that help the analyst

to make well-informed decisions based on a solid foundation of relevant and reliable data. A key advantage of our methodology is that it produces a set of domain-independent threats but, at the same time, it can be tailored towards a set of domain-dependent threats. The SPADA methodology answers the core research question (RQ) stated at the beginning of this article. Precisely, SPADA ensures that the direction pursued by the analyst in modelling security and (soft and hard) privacy threats remains focused and aligned with the desired outcome, as it incorporates the variable elements of threat modelling. These act as guiding principles, allowing the analyst to make informed decisions based on relevant and reliable information.

The specific research questions found an answer in the application of SPADA to the smart car and smart home domains. In particular, SPADA is demonstrated on soft privacy, yielding 23 domain-independent threats, and on hard privacy, yielding 29 domain-independent threats. Each of these lists of threats can be tailored to 43 assets to become smart car domain-dependent (thus answering SRQ1), and to 127 assets to become smart home domain-dependent (thus answering SRQ2). We argue that both soft and hard privacy have not received such an in-depth treatment thus far, considering that our sets of threats are appreciably larger than both LINDDUN's and ENISA's, and they clearly introduce more facets of soft and hard privacy.

Furthermore, the SPADA methodology serves as a valuable tool for risk assessment, as threat modelling lays the groundwork for subsequent risk assessment activities. The process of risk assessment is inherently dynamic, involving continual reassessment of risk levels for acceptance and the implementation of appropriate risk mitigation strategies. In this context, SPADA may practically contribute to dynamicity of risk assessments thanks to a systematic iteration of Step 1, for the setup of its variables, which may be reviewed over the various rounds of the risk assessment activity.

In addition, the SPADA methodology can practically support a DPIA, which is only prescribed by Article 35 of the GDPR in an abstract manner without guidance on implementation. In fact, by contributing to model (soft and hard) privacy threats, SPADA can help organisations in demonstrating their commitment to compliance, in particular to GDPR requirements.

A natural follow-up of our work is to analyse the application of the SPADA methodology to different tuples of variables, for example addressing soft/hard privacy by leveraging internal document sources at a very high level of detail. Regarding the level of detail, we look at the definition of a similarity mapping function to establish a scalable spectrum, rather than limiting it to the two extremes of *Detailed* and *Abstract*. Such a function could be defined, for example, by leveraging an absolute reference point, such as a hypernym, to calibrate the mapping. For any candidate threat or asset whose level of detail should be determined, a semantic similarity measure could be calculated relative to this reference point. Then, the resulting similarity values could be mapped to a linearly ordered range (e.g., from 1 to 5), thereby enabling the analyst to consistently choose their target level of detail in the range, e.g., 2.

Also the elicitation of domain-independent threat starting from domain-dependent threats represents a spark for further investigation, specifically in the pursuit of methods that can automate the extraction of domain-independent information from specific threats.

Moreover, because SPADA uses the embrace operation, which is not fully formalised and/or automated, the SPADA methodology is affected by inherent subjectivity by the analyst, hence this represents a main limitation. Therefore, an investigation on the reduction of subjectivity over the embrace operation represents an interesting direction towards a yet more objective threat modelling methodology. As the execution of SPADA relies on document sources, an inherent limitation stems from the quality of such sources, particularly in terms of the clarity and potential ambiguity of the textual descriptions of the threats and assets they include. Arguably, by augmenting the extracted threats and assets with richer contextual information derived from the same sources (when available), the clarity and precision of the identified elements can be improved, thereby minimising potential textual ambiguities of the identified threats and assets. Hence, a future direction involves the use of Large Language Models (LLMs) to address this challenge, given their promising capabilities in text manipulation, ultimately improving the overall robustness of our methodology. In fact, our future work looks at the implementation of an automated tool for the SPADA steps. Such a tool would not only support the execution of the methodology from scratch, but also streamline subsequent iterations, thus enabling efficient updates to the lists of threats and assets—tasks that are otherwise time-intensive in the manual process. Moreover, such automation would allow for cross-validation between manual and automated results, thereby extending the current partial validation.

An interpretation of the findings of this article is that (soft and hard) privacy is finally threat-modelled as extensively as it deserves and, in particular, as cybersecurity traditionally has been. In particular, the risks for "*natural persons with regard to the processing of personal data and on the free movement of such data*" [23], especially when those natural persons drive smart cars or inhabit smarts, can be now assessed much more precisely than before.

Appendix A Domain-dependent threat elicitation

A.1 Smart car—hard privacy

We present an exemplification of some noteworthy domain-dependent threats, with the additional aim of providing the rationale behind the related threat-asset(s) associations:

$dit\_sc\_h_{i_1}$— *Linked data—unique identifier* refers to the linkage of assets using a unique identifier. Such a linkage poses privacy risks, as it enables tracking and profiling of individuals or vehicles. Affected assets are standard sensors, sensors for autonomous vehicles, telematics box, vehicle ITS station, in-vehicle infotainment (IVI), OBD-II port, back-end systems, database servers, maps servers, third-party service providers servers, sensors data, map data, V2X information, device information, vehicle information, and vehicle maintenance data.

$dit\_sc\_h_{i_2}$— *Combining data of different individuals* specifically impacts user information. The fusion of data from different individuals might lead to privacy breaches and identification risks.

$dit\_sc\_h_{i_3}$— *Profiling an individual through (dis)similarity* involves profiling an individual based on the similarity or dissimilarity of their data. The assets affected by this threat include sensors data, map data, V2X information, device information, user information, special categories of personal data, driver's behaviour, user preferences, vehicle information. By analysing patterns and characteristics in an individual's data, it is possible to create a profile that reveals detailed information about them, their preferences, and their behaviour.

$dit\_sc\_h_{i_4}$— *Identified information* refers to the exposure of identified information, which directly identifies an individual. The assets affected by this threat include sensors data, keys and certificates, map data, V2X information, device information, user information, vehicle information. Identified information, when compromised, can directly link personal data to an individual, leading to privacy risks.

$dit\_sc\_h_{i_5}$— *Pseudonym* involves assets such as sensors data, keys and certificates, map data, V2X information, device information, user information, and special categories of personal data. Pseudonymisation replaces direct identifiers with pseudonyms, but if these pseudonyms can be linked back to individuals, it poses a privacy risk.

$dit\_sc\_h_{i_6}$— *Revealing attributes* involves assets such as sensors data, keys and certificates, map data, V2X information, device information, user preferences, purchase information, vehicle information, and vehicle maintenance data. Certain attributes or data points may inadvertently reveal sensitive information, even if other identifiers are not present.

$dit\_sc\_h_{i_7}$— *The data subject is distinguishable from others* affects user information, special categories of personal data, driver's behaviour, and user preferences. This threat occurs when data contains attributes that distinguish an individual from others in a group, making them easily identifiable.

$dit\_sc\_h_{i_8}$— *Attributable data evidence–Metadata* involves assets such as sensors data, map data, V2X information, driver's behaviour, user preferences, and vehicle maintenance data. Metadata containing identifiable information can inadvertently disclose private details about users or vehicles.

$dit\_sc\_h_{i_9}$— *Observed communications* involves the observation of communications and affects assets such as communication components, in-vehicle infotainment (IVI), OBD-II port, in-vehicle networks, back-end systems, and third-party service providers servers. By intercepting or monitoring communications, adversaries can potentially gain access to sensitive information transmitted between various components or systems.

$dit\_sc\_h_{i_{10}}$— *Insufficient data quality* involves the use of outdated, incorrect or bogus user data, including failure to update or correct the data. The assets affected by this threat include sensors data, keys and certificates, map data, V2X information, device information, user information, special categories of personal data, driver's behaviour, user preferences, purchase information, vehicle information, and vehicle maintenance data.

As an outcome of this exemplification, the resulting number of domain-dependent threats would be:

$$\begin{aligned} &\mathit{affected\_assets}(dit\_sc\_h_{i_1}) + \ldots + \\ &+ \mathit{affected\_assets}(dit\_sc\_h_{i_{10}}) = 16 + 1 + 9 + \\ &+ 7 + 7 + 9 + 4 + 6 + 12 = 71 \end{aligned}$$

A.2 Smart home—soft privacy

We present an exemplification of some noteworthy domain-dependent threats, with the additional aim of providing the rationale behind the related threat-asset(s) associations:

$dit\_sh\_s_{i_1}$— *Unawareness as data subject* describes the situation where individuals are unaware that their personal data is being collected and processed. This threat affects assets such as access and payment credentials for external accounts, users preferences, passwords, user identification, user biometrics, A/V media, documents, pictures, medical, location tracking information, and profile data.

$dit\_sh\_s_{i_2}$— *Lack of data subject control—preferences* focuses on the inability of individuals to control their own preferences. This threat primarily affects users preferences.

$dit\_sh\_s_{i_3}$— *Lack of data subject control—access* refers to individuals being unable to control who has access to their personal data. This threat impacts digital rights management, access and payment credentials for external accounts, smart home setup/structure/inventory information, smart home status information, users preferences, passwords, A/V media, documents, pictures, medical, logs information, location tracking information, and profile data.

$dit\_sh\_s_{i_4}$— *Lack of data subject control—rectification/erasure* involves individuals lacking control over rectifying or erasing their personal data. This threat affects digital rights management, access and payment credentials for external accounts, smart home setup/structure/inventory information, smart home status information, users preferences, passwords, user identification, user biometrics, medical, logs information, location tracking information, and profile data.

$dit\_sh\_s_{i_5}$— *Insufficient data subject controls* indicates inadequate mechanisms for users to control their personal data. Similarly to the previous case, this threat affects digital rights management, access and payment credentials for external accounts, smart home setup/structure/inventory information, smart home status information, users preferences, passwords, user identification, behavioural patterns and trends, user biometrics, medical, logs information, location tracking information, and profile data.

$dit\_sh\_s_{i_6}$— *Invalid consent* pertains to obtaining consent that is not legally valid. This threat affects all assets within the smart home environment.

$dit\_sh\_s_{i_7}$— *Violation of storage limitation principle* involves retaining personal data for longer than necessary. This threat impacts access and payment credentials for external accounts, smart home setup/structure/inventory information, smart home status information, users preferences, passwords, user identification, user biometrics, medical, logs information, location tracking information, and profile data.

$dit\_sh\_s_{i_8}$— *Insufficient cybersecurity risk management* indicates a lack of adequate management of cybersecurity risks. This threat primarily affects security management and operation within the smart home environment.

$dit\_sh\_s_{i_9}$— *Unauthorized use of IPR protected resources* involves using intellectual property rights (IPR) protected resources without authorisation. This threat affects Value/IPRs within the smart home environment.

$dit\_sh\_s_{i_{10}}$— *Judiciary decisions/court orders* pertains to legal decisions or orders that affect privacy and data handling. This threat impacts all assets within the smart home environment.

As an outcome of this exemplification, the resulting number of domain-dependent threats would be:

$$\begin{aligned} &affected\_assets(dit\_sh\_s_{i_1}) + \ldots + \\ &+ affected\_assets(dit\_sh\_s_{i_{10}}) = 11 + 1 + 13 + \\ &+ 12 + 13 + 127 + 11 + 1 + 1 + 127 = 317 \end{aligned}$$

A.3 Smart home—hard privacy

we present an exemplification of some noteworthy domain-dependent threats, with the additional aim of providing the rationale behind the related threat-asset(s) associations:

$dit\_sh\_h_{i_1}$— *Linkable data—through combination* involves linking data items through combinations. This threat impacts access and payment credential for external accounts, user identification, behavioural patterns and trends, user biometrics, medical, logs information, location tracking information, and profile data.

$dit\_sh\_h_{i_2}$— *Linkable data—through profiling, derivation, or inference* entails linking data items through profiling, derivation, or inference. This threat affects access and payment credential for external accounts, smart home

setup/structure/ inventory information, smart home status information, users preferences, user identification, behavioural patterns and trends, user biometrics, medical, logs information, location tracking information, and profile data.

$dit\_sh\_h_{i_3}$— *Profiling a group of individuals* involves profiling a group of individuals. Similarly to the previous case, this threat impacts access and payment credential for external accounts, smart home setup/structure/inventory information, smart home status information, users preferences, user identification, behavioural patterns and trends, user biometrics, medical, logs information, location tracking information, and profile data.

$dit\_sh\_h_{i_4}$— *Processing of identified information* refers to processing, in a broader-GDPR-like sense, of identified information for various purposes. This threat impacts billing, ordering, digital rights management, identification, authentication, security, trouble shooting/diagnosis, service personnel access, access and payment credential for external accounts, user identification, user biometrics, medical, connected transportation, specialised terminal, smart phone, tablet computer, desktop computer/pc, and profile data.

$dit\_sh\_h_{i_5}$— *Identified information in metadata* involves the presence of identified data present in metadata. This threat affects access and payment credential for external accounts, medical, and profile data.

$dit\_sh\_h_{i_6}$— *Identifiable information* pertains to information that can be easily identified. This threat impacts access and payment credential for external accounts, user identification, user biometrics, medical, and profile data.

$dit\_sh\_h_{i_7}$— *Pseudonym–identifier* involves the use of pseudonyms linked to identifiers. This threat impacts access and payment credential for external accounts, user identification, user biometrics, medical, and profile data.

$dit\_sh\_h_{i_8}$— *Attributable data evidence* pertains to evidence that attributes data to a specific individual. This threat affects access and payment credential for external accounts, smart home setup/structure/inventory information, smart home status information, users preferences, passwords, user identification, user biometrics, A/V media, documents, pictures, medical, external cloud storage, network attached storage, removable media, logs information, location tracking information, and profile data.

$dit\_sh\_h_{i_9}$— *Application side-effect* refers to the detectability of relevant information by side-effects caused by applications. This threat impacts operating system(s), device drivers, applications, firmware, specialised terminal, interface to home gateway, and cloud server.

$dit\_sh\_h_{i_{10}}$— *System responses* relates to the (malicious) analysis of responses generated by systems. This threat affects billing, ordering, digital rights management, identification, authentication, security, trouble shooting/diagnosis, service personnel access, updates, connected transportation, vacuum cleaner, lawn mower, mobile robotics telepresence, temperature, light, microphones, humidity/moisture, gas/smoke/CO2 detectors, motion, face recognition/biometrics, electrical current/on-off, door (magnet), lock, physiological sensor, wearable, specialised terminal, switch, router, bridge, repeater, modem, gateway, firewall, smart TV, displays, speakers, digital photo frame, refrigerator, washing machine, dish washer, oven, humidifier/dehumidifier, food processor, drinks makers, and cloud server.

As an outcome of this exemplification, the resulting number of domain-dependent threats would be:

$$\begin{aligned} &affected\_assets(dit\_sh\_h_{i_1}) + \ldots + \\ &+ affected\_assets(dit\_sh\_h_{i_{10}}) = 8 + 11 + 11 + \\ &+ 18 + 3 + 5 + 5 + 17 + 7 + 45 = 130 \end{aligned}$$

Appendix B Partial validation

B.1 Partial validation of smart car—hard privacy

The Hacker News [40] reported that multiple bugs affecting millions of vehicles from 16 different manufacturers could be abused to unlock, start, and track cars, plus impact the privacy of car owners. The vulnerabilities were found in the automotive APIs powering several brands, including, among the others, BMW, Ferrari, Mercedes and Toyota. The article mentions vulnerabilities that could permit user account takeover and the disclosure of sensitive information, indicating the potential for attributes to be revealed. This falls under the "Revealing attributes" threat from Table 2. Furthermore, the vulnerabilities found in connected vehicle services, such as those provided by SiriusXM and Spireon, could potentially allow for remote attacks and the ability to issue arbitrary com-

mands. This indicates the potential for detecting involvement through "Observed communications".

In addition, CPO Magazine [36] revealed that personal information of over 2 million Aflac life insurance and Zurich auto insurance policyholders in Japan was leaked online in a third-party data breach. If we consider the insurance component within the entire automotive ecosystem, which is thus not limited to smart cars only, this breach indicates the potential for linking data items (policyholder information) to gain unauthorised access and potentially track individuals, thus resulting into the "Linked data" threat.

Moreover, an article by BleepingComputer [5] mentions a data breach at Toyota Motor Corporation where the car-location information of 2,150,000 customers was exposed. The breach resulted from a misconfiguration of the cloud environment, allowing unauthorised access to the data. While the exposed information includes in-vehicle GPS navigation terminal ID numbers, chassis numbers, and vehicle location information with time data, it is stated that this does not constitute personally identifiable information (PII) unless the attacker knows the vehicle identification number (VIN). Nonetheless, the exposure of this information still raises concerns about potential identifying threats, specifically "Identified information", if combined with other data sources.

B.2 Partial validation of smart home—soft privacy

Threatpost [57] reported that researchers from Cisco Talos have discovered 20 vulnerabilities in Samsung's SmartThings Hub, a centralised controller used to manage various IoT devices. These vulnerabilities could potentially allow attackers to gain unauthorised access to the connected devices, control them remotely, and perform malicious activities. The vulnerabilities are located in the SmartThings Hub's Linux-based firmware, which communicates with IoT devices via Zigbee, Z-Wave, and Bluetooth protocols. In particular, the compromised devices could include smart locks, cameras, motion detectors, and thermostats, among others. This incident may reflect the need to address an "Insufficient cybersecurity risk management", which may be related to privacy concerns along with the cybersecurity aspects.

In addition, CNET [11] reported that Amazon was fined by the Federal Trade Commission (FTC) for privacy violations related to its smart home products, specifically Ring and Alexa. Amazon was asked to pay $25 million for not deleting children's data collected through Alexa and $5.8 million for failing to restrict employee and contractor access to Ring security videos. The FTC alleged that Amazon stored children's voice and geolocation data acquired through Alexa, using it to improve the Alexa algorithm, which put the data at risk. Furthermore, Ring, a company owned by Amazon, was penalised for allegedly allowing employees and contractors to access customer videos without proper consent, and for not implementing adequate security measures. The fines will be used to refund customers, and both Amazon and Ring have stated their disagreements with the FTC's claims. Such piece of news encompasses multiple soft privacy threats that we elicited in our demonstration, i.e., "Unawareness of processing", "Lack of data subject control", and "Violation of data minimization principle".

B.3 Partial validation of smart home—hard privacy

Forbes [22] reported a breach involving a Chinese company called Orvibo, which manages an IoT platform for smart home devices. The company's user database, containing over 2 billion logs, was left exposed on the internet without password protection. The database included sensitive information such as email addresses, passwords, account reset codes, geolocation, IP addresses, and more. This incident, apart from traditional cybersecurity threats, involved "Linked data" and/or "Linkable data" threats as the breach exposed a variety of data, including user passwords, preferences, account reset codes, geolocation, and more, that could be potentially linked/linkable. Furthermore, it appears clear that the breach also provided an "Attributable data evidence" threat.

Moreover, researchers from Italy and the UK have discovered a vulnerability [6] in TP-Link smart bulbs that could be exploited by cybercriminals to gather Wi-Fi credentials. The researchers used Vulnerability Assessment and Penetration Testing (VAPT), specifically following the PETIoT IoT-focused Kill Chain [3], to assess the security of the Tp-Link Tapo Smart Wi-Fi Multicolor Light Bulb (L530E). The threats arising from such vulnerability mainly concern cybersecurity as a target property, yet some of the hard privacy threats that we elicited may be involved as well, i.e., "Linked data" and/or "Linkable data". The adoption of PETIoT also led to a breach of motion detection on the TP-Link TAPO C200, namely the best-selling IP camera on Amazon Italy. It means that a threat agent can infer whether the camera detects motion despite the fact that the corresponding notifications are encrypted. This naturally materialises into the "Attributable action side-effect evidence" and "Observed communications" hard privacy threats.

Acknowledgements We thank Dimitri Van Landuyt for the precious discussions and suggestions. This work acknowledges financial support from: Italian PNRR 2022 SERICS Spoke 6, Task 1.2, Project "SCAI—Supply Chain Attack Avoidance".

Funding Open access funding provided by Scuola IMT Alti Studi Lucca within the CRUI-CARE Agreement.

Data Availability All data are available online in a repository [48].

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence this work.

References

1. Ansari, T.J., Pandey, D.: An integration of threat modeling with attack pattern and misuse case for effective security requirement elicitation. Int. J. Adv. Res. Comput. Sci. **8**, 16–20 (2017). (https://api.semanticscholar.org/CorpusID:86796612)
2. Anwar, M.N., Nazir, M., Mustafa, K. Security threats taxonomy: smart-home perspective. In: 2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall), pp. 1–4, (2017). https://doi.org/10.1109/ICACCAF.2017.8344666
3. Bella, G., Biondi, P., Bognanni, S., et al.: Petiot: penetration testing the internet of things. Internet Things **22**, 100707 (2023). https://doi.org/10.1016/j.iot.2023.100707. (https://www.sciencedirect.com/science/article/pii/S2542660523000306)
4. Bella, G., Biondi, P., Tudisco, G.: A double assessment of privacy risks aboard top-selling cars. Autom. Innov. –18. (2023b). https://doi.org/10.1007/s42154-022-00203-2
5. BleepingComputer: Toyota: car location data of 2 million customers exposed for ten years. (2023a). https://www.bleepingcomputer.com/news/security/toyota-car-location-data-of-2-million-customers-exposed-for-ten-years/
6. BleepingComputer: TP-Link smart bulbs can let hackers steal your WiFi password. (2023b). https://www.bleepingcomputer.com/news/security/tp-link-smart-bulbs-can-let-hackers-steal-your-wifi-password/
7. Brous, P., Janssen, M., Herder, P.: The dual effects of the internet of things (IoT): a systematic review of the benefits and risks of IoT adoption by organizations. Int. J. Inf. Manage. **51**, 101952 (2020). https://doi.org/10.1016/j.ijinfomgt.2019.05.008. (https://www.sciencedirect.com/science/article/pii/S0268401218309022)
8. Buil-Gil, D., Kemp, S., Kuenzel, S., et al.: The digital harms of smart home devices: a systematic literature review. Comput. Hum. Behav. **145**, 107770 (2023). https://doi.org/10.1016/j.chb.2023.107770. (https://www.sciencedirect.com/science/article/pii/S0747563223001218)
9. Canedo, E.D., Bandeira, I.N., Calazans, A.T.S., et al.: Privacy requirements elicitation: a systematic literature review and perception analysis of it practitioners. Requir. Eng. **28**, 177–194 (2022). (https://api.semanticscholar.org/CorpusID:258570284)
10. Chah, B., Lombard, A., Bkakria, A., et al.: Privacy threat analysis for connected and autonomous vehicles. Procedia Comput. Sci. **210**, 36–44. (2022). https://doi.org/10.1016/j.procs.2022.10.117, https://www.sciencedirect.com/science/article/pii/S1877050922015733, the 13th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN) / The 12th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2022) / Affiliated Workshops
11. CNET: Amazon to pay $30M for ring and alexa privacy violations: tips for protecting your smart home data. (2023). https://www.cnet.com/tech/services-and-software/amazon-to-pay-30-m-for-ring-and-alexa-privacy-violations-tips-for-protecting-your-smart-home-data/
12. Cybernews: Bmw exposes clients in italy. (2023). https://cybernews.com/security/bmw-exposes-italy-clients/
13. Danezis, G.: Introduction to privacy technology. (2008). http://www0.cs.ucl.ac.uk/staff/G.Danezis/talks/Privacy_Technology_cosic.pdf
14. Danezis, G., Gurses, S.: A critical review of 10 years of privacy technology. In: Proceedings of surveillance cultures: a global surveillance society, pp. 1–16, (2010). https://www.researchgate.net/publication/228538295_A_critical_review_of_10_years_of_Privacy_Technology
15. Deloitte: 2023 global automotive consumer study. (2023). https://www.deloitte.com/global/en/Industries/automotive/perspectives/global-automotive-consumer-study.html
16. Deng, M., Wuyts, K., Scandariato, R., et al.: A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Requir. Eng. **16**(1), 3–32 (2011). https://doi.org/10.1007/s00766-010-0115-7
17. ENISA: threat landscape and good practice guide for smart home and converged media. (2014). https://www.enisa.europa.eu/publications/threat-landscape-for-smart-home-and-media-convergence
18. ENISA. Threat taxonomy. (2016). https://www.enisa.europa.eu/topics/cyber-threats/threats-and-trends/enisa-threat-landscape/threat-taxonomy/view
19. ENISA. Good practices for security of smart cars. (2019). https://www.enisa.europa.eu/publications/smart-cars
20. ENISA. ENISA threat landscape 2022. (2022). https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022
21. Europe, A.N.: Tesla escapes fine from dutch watchdog after automaker alters security cameras. (2023). https://europe.autonews.com/automakers/tesla-alters-cameras-avoid-dutch-fine-over-privacy-violations
22. Forbes: confirmed: 2 billion records exposed in massive smart home device breach. (2019). https://www.forbes.com/sites/daveywinder/2019/07/02/confirmed-2-billion-records-exposed-in-massive-smart-home-device-breach/
23. GDPR: Regulation (EU) 2016/679 general data protection regulation. (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679
24. Geneiatakis, D., Kounelis, I., Neisse, R., et al.: Security and privacy issues for an IoT based smart home. In: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1292–1297, (2017). https://doi.org/10.23919/MIPRO.2017.7973622
25. Ghirardello, K., Maple, C., Ng, D., et al.: Cyber security of smart homes: development of a reference architecture for attack surface analysis. In: Living in the Internet of Things: Cybersecurity of the IoT–2018, pp. 1–10 (2018). https://doi.org/10.1049/cp.2018.0045
26. Heartfield, R., Loukas, G., Budimir, S., et al.: A taxonomy of cyber-physical threats and impact in the smart home. Comput. Secur. **78**, 398–428 (2018). https://doi.org/10.1016/j.cose.2018.07.011. (https://www.sciencedirect.com/science/article/pii/S0167404818304875)
27. Henriques de Gusmão, A.P., Mendonça Silva, M., Poleto, T., et al.: Cybersecurity risk analysis model using fault

tree analysis and fuzzy decision theory. Int. J. Inf. Manage. **43**, 248–260 (2018). https://doi.org/10.1016/j.ijinfomgt.2018.08.008. (https://www.sciencedirect.com/science/article/pii/S026840121830077X)
28. Hevner, A., March, S., Park, J., et al.: Design science in information systems research. Manag. Inf. Syst. Q. **28**, 75 (2004)
29. Hoepman, J.H.: Privacy is hard and seven other myths: achieving privacy through careful design. (2021). https://doi.org/10.7551/mitpress/12587.001.0001
30. Howard, M., Lipner, S.: The Security Development Lifecycle, vol 34. Microsoft Press (2006). https://doi.org/10.1007/s11623-010-0021-7
31. Islam, S., Mouratidis, H., Wagner, S.: Towards a framework to elicit and manage security and privacy requirements from laws and regulations. In: Requirements Engineering: Foundation for Software Quality, (2010). https://api.semanticscholar.org/CorpusID:15084158
32. ISO: ISO/IEC 21434-road vehicles-cybersecurity engineering (2021)
33. Kavallieratos, G., Chowdhury, N., Katsikas, S., et al.: Threat analysis for smart homes. Future Internet 11(10). (2019a). https://doi.org/10.3390/fi11100207, https://www.mdpi.com/1999-5903/11/10/207
34. Kavallieratos, G., Gkioulos, V., Katsikas, S.K.: Threat analysis in dynamic environments: the case of the smart home. In: 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), pp. 234–240, (2019b). https://doi.org/10.1109/DCOSS.2019.00060
35. Law, B.: New us agency joins fray over Massachusetts repair law, car data. (2023) https://news.bloomberglaw.com/privacy-and-data-security/new-us-agency-joins-fray-over-massachusetts-repair-law-car-data
36. Magazine, C.: Data breach exposed 2 million Aflac and Zurich insurance policyholders- records. (2023). https://www.cpomagazine.com/cyber-security/data-breach-exposed-2-million-aflac-and-zurich-insurance-policyholders-records/
37. Makri, E.L., Lambrinoudakis, C.: Privacy principles: towards a common privacy audit methodology. In: Fischer-Hübner, S., Lambrinoudakis, C., López, J. (eds.) Trust, Privacy and Security in Digital Business, pp. 219–234. Springer International Publishing, Cham (2015)
38. Microsoft: the STRIDE threat model. (2009). https://learn.microsoft.com/en-us/previous-versions/commerce-server/ee823878%28v=cs.20,
39. Myagmar, S., Lee, A.J., Yurcik, W.: Threat modeling as a basis for security requirements. (2005). https://api.semanticscholar.org/CorpusID:9164059
40. News, T.H.: Millions of vehicles at risk: api vulnerabilities uncovered in 16 major car brands. (2023). https://thehackernews.com/2023/01/millions-of-vehicles-at-risk-api.html
41. NIST: foundations of a security policy for use of the national research and educational network (NISTIR 4734). (1992). https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir4734.pdf
42. OWASP: top 10 privacy risks. (2021). https://owasp.org/www-project-top-10-privacy-risks/
43. OWASP: threat modeling cheat sheet. (2019). https://cheatsheetseries.owasp.org/cheatsheets/Threat_Modeling_Cheat_Sheet.html
44. Pattakou, A., Kalloniatis, C., Gritzalis, S.: Security and privacy requirements engineering methods for traditional and cloud-based systems: a review. In: Proceedings of the 8th International Conference on Cloud Computing, GRIDs, and Virtualization, (2017). https://api.semanticscholar.org/CorpusID:252048830
45. Peffers, K., Tuunanen, T., Rothenberger, M., et al.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**, 45–77 (2007)
46. Pfitzmann, A., Hansen, M. (2010).: A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf, v0.34
47. Pompigna, A., Mauro, R.: Smart roads: a state of the art of highways innovations in the smart age. Eng. Sci. Technol. Int. J. **25**, 100986 (2022). https://doi.org/10.1016/j.jestch.2021.04.005. (https://www.sciencedirect.com/science/article/pii/S2215098621000872)
48. Raciti, M., Bella, G.: Github repository with complete outcomes. (2023a). https://github.com/tsumarios/Privacy-Threat-Modelling-Research/tree/main/SPADA
49. Raciti, M., Bella, G.: How to model privacy threats in the automotive domain. In: Proceedings of the 9th International Conference on Vehicle Technology and Intelligent Transport Systems–VEHITS, INSTICC. SciTePress, pp. 394–401, (2023b). https://doi.org/10.5220/0011998800003479
50. Raciti, M., Bella, G.: A threat model for soft privacy on smart cars. In: In 2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)., (2023c). https://doi.org/10.1109/EuroSPW59978.2023.00005
51. Raciti, M., Bella, G.: Up-to-date threat modelling for soft privacy on smart cars. In: Katsikas, S., Cuppens, F., Cuppens-Boulahia, N., et al. (eds) Computer Security. ESORICS 2023 International Workshops. Springer Nature Switzerland, Cham, pp. 454–473, (2024). https://doi.org/10.1007/978-3-031-54204-6_27
52. Reuters: Toyota's Indian unit warns of a possible customer data breach. (2023). https://www.reuters.com/technology/toyotas-indian-unit-warns-possible-customer-data-breach-2023-01-01/
53. SAE: taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (J3016_202104). (2021). https://www.sae.org/standards/content/j3016_201806/
54. Schneier, B.: Attack trees. Dr Dobb's J. **24**(12), 21–29 (1999)
55. Siwakoti, Y.R., Bhurtel, M., Rawat, D.B., et al.: Advances in IoT security: vulnerabilities, enabled criminal services, attacks, and countermeasures. IEEE Internet Things J. **10**(13), 11224–11239 (2023). https://doi.org/10.1109/JIOT.2023.3252594
56. TechCrunch: Toyota japan exposed millions of vehicles' location data for a decade. (2023). https://techcrunch.com/2023/05/12/toyota-japan-exposed-millions-locations-videos/
57. Threatpost: bugs in samsung IoT Hub leave smart home open to attack. (2018). https://threatpost.com/bugs-in-samsung-iot-hub-leave-smart-home-open-to-attack/134454/
58. Toh CSanguesa, M.: Advances in smart roads for future smart cities. In: Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, (2020). https://doi.org/10.1098/rspa.2019.0439
59. Underscored, C.: The ring car cam takes ring's great security smarts on the road. (2023). https://edition.cnn.com/cnn-underscored/reviews/ring-car-cam
60. Van Landuyt, D., Joosen, W.: A descriptive study of assumptions made in linddun privacy threat elicitation. In: Proceedings of the 35th Annual ACM Symposium on Applied Computing. Association for Computing Machinery, New York, NY, USA, SAC '20, pp. 1280–1287, (2020). https://doi.org/10.1145/3341105.3375762,
61. Vasenev, A., Stahl, F., Hamazaryan, H., et al.: Practical security and privacy threat analysis in the automotive domain: long term support scenario for over-the-air updates. In: Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems–VEHITS,, INSTICC. SciTePress, pp. 550–555, (2019). https://doi.org/10.5220/0007764205500555
62. Wang, Y., Wang, Y., Qin, H., et al.: A systematic risk assessment framework of automotive cybersecurity. Automot. Innov. **4**(3), 253–261 (2021). https://doi.org/10.1007/s42154-021-00140-6

63. Wikipedia: Semantic Relations. (2023). https://en.wiktionary.org/wiki/Wiktionary:Semantic_relations
64. Wuyts, K., Joosen, W.: Linddun privacy threat modeling: a tutorial. Technical Report (CW Reports), (2015)
65. Wuyts, K., Van Landuyt, D., Hovsepyan, A., et al.: Effective and efficient privacy threat modeling through domain refinements. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. Association for Computing Machinery, New York, NY, USA, SAC '18, pp. 1175–1178, (2018). https://doi.org/10.1145/3167132.3167414,
66. Wuyts, K., Sion, L., Van Landuyt, D., et al.: Knowledge is power: systematic reuse of privacy knowledge for threat elicitation. In: 2019 IEEE Security and Privacy Workshops (SPW), pp. 80–83, (2019). https://doi.org/10.1109/SPW.2019.00025
67. Ziegeldorf, J.H., Morchon, O.G., Wehrle, K.: Privacy in the internet of things: threats and challenges. Secur. Commun. Netw. **7**(12), 2728–2742 (2014). https://doi.org/10.1002/sec.795